Overdiagnosis and Undertesting for Infectious Diseases*

Tinglong Dai Shubhranshu Singh

Marketing Science, Forthcoming

April 2024

Abstract

The COVID-19 pandemic brought the availability of diagnostic tests to the forefront of public attention and highlighted the prevalence of undertesting (i.e., insufficient test supply relative to demand). Another important yet little studied systematic issue is overdiagnosis (i.e., positive diagnoses for patients with negligible viral loads): evidence suggests U.S. laboratories have adopted highly sensitive diagnosis criteria, such that up to an estimated 90% of positive diagnoses are for minuscule viral loads. Motivated by this situation, we develop a theory of diagnostic testing for infectious diseases that explains both undertesting and overdiagnosis. We show a commercial laboratory has an incentive to inflate its diagnosis criterion, which generates a higher diagnosis-driven demand as a result of contact-tracing efforts, albeit while dampening demand from disease transmission. An inflated diagnosis criterion prompts the laboratory to build a higher testing capacity, which may not fully absorb the inflated demand, so undertesting arises. Finally, we examine a social planner's problem of whether to mandate that the laboratory report the viral load along with its diagnosis, so that a physician or contact tracer can make informed triage decisions. Our results show the social planner may choose *not* to mandate viral-load reporting initially; this choice induces a higher testing capacity and can help reduce disease transmission.

*Both authors contributed equally to this work and are listed alphabetically. Address: Carey Business School, Johns Hopkins University, Baltimore, MD 21202. Email: dai@jhu.edu, shubhranshu.singh@jhu.edu. We are grateful to the Senior Editor, Associate Editor, and two anonymous reviewers for the constructive review process. We thank Anthony Dukes, Michael Fu, Guodong (Gordon) Gao, Erica Gralla, Kathy McDonald, and Sridhar Seshadri for their feedback and guidance. We also thank seminar participants at Arizona State University, Chinese University of Hong Kong, University of Illinois Urbana-Champaign, and University of Maryland, College Park. An earlier version of the paper was selected as the lead article of *COVID Economics* (Issue 58) and featured in *VoxEU*. The *New York Times* featured another earlier version of the paper in a January 6, 2022 story.

1. Introduction

An important theme of the literature on diagnostic services (e.g., Arora and Fosfuri 2005; Dai and Singh 2020; Mullainathan and Obermeyer 2022; Sarvary 2002) is the provision of testing. Whereas much of the literature focuses on the phenomenon of overprovision by diagnostic experts (McGuire 2000), the coronavirus disease 2019 (COVID-19) pandemic reveals the prevalence of undertesting (i.e., a shortage of laboratory testing capacity relative to the demand for testing): during the initial months of the pandemic, the U.S. failed to roll out abundant and usable testing, resulting in widespread community infection that was not contained or detected (Shear et al. 2020). For much of the pandemic, the U.S. continued to experience severe undertesting during periods of high demand (Rosenthal 2021).

Whereas the issue of undertesting during the pandemic has been widely recognized, another important issue, overdiagnosis, has not received due attention. By "overdiagnosis," we do not mean patients are given wrong diagnoses. Rather, we mean patients with negligible viral loads are given positive diagnoses.[1] According to a *New York Times* article entitled "Your Coronavirus Test is Positive. Maybe It Shouldn't Be" (Mandavilli 2020a),

> In three sets of testing data that include cycle thresholds, compiled by officials in Massachusetts, New York and Nevada, up to 90 percent of people testing positive carried *barely any* virus [emphasis added], a review by The Times found. On Thursday, the United States recorded 45,604 new coronavirus cases, according to a database maintained by The Times. If the rates of contagiousness in Massachusetts and New York were to apply nationwide, then perhaps only 4,500 of those people may actually need to isolate and submit to contact tracing.

Overdiagnosis arises because when conducting the polymerase chain reaction (PCR) test for

[1] The concept of overdiagnosis has been widely used in medicine. In the context of cancer, Esserman et al. (2013) define overdiagnosis as a scenario "when tumors are detected that, if left unattended, would not become clinically apparent or cause death." Welch et al. (2011) use the analogy of a turtle to illustrate that certain cancers progress so slowly that they are unlikely to have meaningful health impacts during one's lifetime. More broadly, Brodersen et al. (2018) define overdiagnosis as "making people patients unnecessarily, by identifying problems that were never going to cause harm or by medicalising ordinary life experiences through." Our use of the concept is aligned with the medical literature while incorporating the infection risk that a patient poses to others. Note "overdiagnosis" differs from "overtesting"; the latter refers to overutilization of diagnostic testing (Brodersen et al. 2018).

COVID-19,[2] U.S. commercial laboratories rely on a rather large magnification factor, as measured by the cycle threshold (CT), that is, the number of cycles necessary for spotting the coronavirus. (The PCR process amplifies DNA segments cycle by cycle. A total of 30 cycles results in 2^{30}, approximately 1 billion, copies of the original DNA segments, whereas a total of 40 cycles results in 2^{40}, approximately 1 *trillion*, copies.) A low CT means a high viral load, and vice versa (Mandavilli 2020b). According to Mandavilli (2020a), "most tests set the limit at 40, a few at 37," which means patients are given positive diagnoses if the PCR test takes up to 40 or 37 cycles. Yet, "tests with thresholds so high may detect not just live virus but also genetic fragments, leftovers from infection that pose no particular risk—akin to finding a hair in a room long after a person has left." For this reason, experts agree that tests with a CT above 35 are too sensitive and that "a more reasonable cutoff would be 30 to 35." Echoing the above *New York Times* article, in a July 17, 2020, interview with *This Week in Virology*, Dr. Anthony Fauci explained (TWiV 2020),

> If you get a cycle threshold of 35 or more, the chance of it being replication competent are minuscule... It's very frustrating for the patients as well as for the physicians that someone comes in and they repeat their PCR and it's like 37 cycle threshold, but you almost never can culture virus from a 37, 38, even 36 cycle. It's just dead nucleotides. Period.

Even as U.S. laboratories rely on hypersensitive diagnostic criteria, they rarely, if ever, report the CT values along with a binary diagnosis (i.e., positive or negative), blinding physicians, contact tracers, and others who seek to assess the viral load behind the binary diagnosis.[3] Physician leaders in the U.S. increasingly demand a departure from the lack of reporting of viral load, as evidenced in a recent *Science* article entitled "A Call for Diagnostic Tests to Report Viral Load" (Service 2020), which contends that including the CT value in

[2]We focus on the PCR test throughout the paper. In practice, a number of "rapid tests," such as the BinaxNOW test developed by Abbott Laboratories, have been used in various situations. However, experts have expressed concerns about their relatively low sensitivity and high false-negative rates (Madrigal and Meyer 2020). Another reason we focus on the PCR test is that it is a standard diagnostic tool for many infectious diseases, including Ebola, HIV, SARS, and tuberculosis (Salis 2009), meaning our model has broader implications for future epidemics or pandemics.

[3]Some of the largest COVID testing labs, such as Laboratory Corporation of America Holdings (Labcorp) and Quest Diagnostics Inc., are also PCR test manufacturers themselves; see https://bit.ly/commericalcovidtests for a list of commercially available COVID-19 diagnostic tests. Consistent with this observation, our main model does not separately model a lab and a manufacturer for the test.

the diagnosis can help (1) contact tracers "triage their efforts based on CT values" and (2) physicians "flag patients most at risk for severe disease and death." The call has been echoed outside the U.S. In India, "many doctors are now telling patients that their COVID-19 test reports should mention the cycle threshold (CT) value, and not just the positive or negative outcome" (Rao 2020). Despite these calls for action, "little effort has been made to track viral loads" (Mandavilli 2020b).

In this paper, motivated by the situation, we develop a theory of laboratory testing that explains both undertesting and overdiagnosis amid an epidemic. A commercial laboratory (hereafter "lab") sets the CT cutoff (i.e., the maximum number of cycles before the PCR test stops), which it uses to determine whether a patient is positive or negative: if a patient has a level of virus that is detectable before the number of cycles reaches the CT cutoff, the patient is diagnosed as positive; conversely, if the level is not detectable within those cycles, the patient is diagnosed as negative.[4] Consistent with the standard receiver operating characteristic (ROC) curve approach to determining cutoffs for diagnostic tests (McNeil et al. 1975; Ekelund 2011), in the absence of revenue incentives, the lab would choose a CT cutoff that minimizes the sum of the expected costs to the patient of a false-positive diagnosis and a false-negative diagnosis. This approach assumes maximum diagnostic uncertainty for each patient tested and results in a CT cutoff that does not depend on the prevalence of the disease in the population. In Section 5.1, we allow for the possibility of a prevalence-dependent CT cutoff and show that our main results are robust to the alternative approach of setting the CT cutoff. In addition to choosing the CT cutoff and deciding whether to report the CT value on test results, the lab chooses its future capacity level by weighing demand for testing resulting from both contact tracing and community spread.

Most diagnostic service providers in the U.S., including publicly listed companies such as Quest Diagnostics, are for-profit commercial entities (Foley 2020). At the same time, they are providing an essential public service. Thus, the lab's decision-making is motivated by both profitability and patient well-being. We show that, under reasonable parametric

[4]The literature has documented that determining the optimal CT cutoff is an important decision for diagnostic labs performing PCR tests for various diseases. For example, Caraguel et al. (2011) state, "Diagnostic laboratories frequently select a subjective cutoff value for real-time amplification assays, above which a threshold cycle (Ct) value is deemed false."

assumptions, the lab has an incentive to inflate the PCR diagnosis criterion, as specified by the CT cutoff. The inflated CT cutoff generates a higher demand for testing as a result of contact-tracing efforts. At the same time, the inflated CT cutoff reduces the chance of false-negative diagnoses, which thus helps reduce community spread; as a result, the lab faces dampened demand from disease transmission. In view of this tradeoff, our further analysis reveals the lab is more likely to inflate the diagnosis criterion if the intensity of contact tracing is high relative to the expected contagiousness.

Our model allows the lab's diagnostic policy to interact with its capacity decision. We show an inflated diagnosis criterion incentivizes the lab to build a higher testing capacity. Yet, the increased capacity may not fully absorb the inflated demand. As a result, the phenomenon of undertesting (i.e., the probability that all demand can be satisfied is lower than when the lab does not inflate the diagnosis criterion) can arise despite the high testing capacity. In this sense, we have developed a unifying theory that explains both undertesting and overdiagnosis of infectious diseases.

We also analyze a social planner's regulatory problem. In the U.S., a lab's diagnostic testing process and interpretation of testing results are considered part of the practice of medicine and cannot be directly regulated (Shirts 2020). Despite the social planner's inability to control the sensitivity of diagnostic tests, a possible regulatory decision entails whether to mandate the lab to report the viral-load information along with a binary diagnosis, such that a physician or contact tracer can better assess the patient's condition (Mandavilli 2020b). We show not mandating the lab's reporting of viral-load information may be in the social planner's best interest for the following reason: When the lab is mandated to report the viral-load information along with each test result, other parties (including physicians, contact tracers, and patients) can use the viral-load information to triage contact-tracing efforts that contribute to additional demand for testing. Stated differently, by disclosing the viral-load information, the lab essentially gives up its demand-inducing power to others. Because other parties will rely on a lower CT cutoff in their decisions, the lab faces a lower total demand, all else being the same. Accordingly, the lab does not elevate testing capacity to a level that provides the same level of testing availability. In addition, a lower CT cutoff increases the patient's likelihood of receiving a false-negative test result, which contributes to

community spread. Therefore, in light of public interest, the social planner may find that not mandating the reporting of the viral-load information initially is more desirable. This result highlights the tensions between the objectives of various stakeholders and sheds light on delicate tradeoffs that policymakers have to make amid an epidemic or pandemic.

Our paper contributes to the ongoing discourse on diagnostic testing by providing a novel theory that explains both overdiagnosis and undertesting. Although our findings are broadly consistent with empirical observations, we do not claim our theory is the sole explanation for these phenomena or that it captures all the forces driving a lab's CT cutoff and capacity decisions; we discuss several alternative explanations in Section 6. Our analysis is from a descriptive perspective and suggests that in the absence of direct public investment in testing capacity, both phenomena, as undesirable as they are, may necessarily arise and are indeed a result of the social planner's rational choice. These results have important implications for the ongoing and future epidemics. Nobel laureate Gary Becker (2020) estimates, "The expected worldwide cost in terms of willingness to pay to avoid the risk of another great pandemic that had a one in hundred probability of occurring during the next twenty years would be approximately 1/100 * $20 trillion, or about $200 billion. This cost would justify sizable increases in [public health investments]" (p. 55). In a *JAMA* article, Cutler and Summers (2020) estimate the total cost of the COVID-19 pandemic to the U.S. at $16 trillion, more than 90% of the country's annual economic output. In light of the magnitude of these estimates, our findings have implications for countries that have built testing capacity primarily through the private sector. Public-private partnerships and innovative contracting mechanisms aimed at disentangling labs' diagnostic testing decisions from incentives to build testing capacity may be worth exploring.

The rest of the paper is organized as follows. In Section 2, we discuss how our paper relates and contributes to the relevant literature in economics and marketing. Section 3 sets up our modeling framework. Section 4 analyzes our model and generates managerial implications related to the lab's diagnostic threshold and capacity decisions, as well as policy implications related to overdiagnosis and undertesting. Section 5 presents an alternative theory of overdiagnosis and two extensions of the main model. Section 6 concludes.

2. Literature

Our paper is related to the literature on diagnostic services. This literature originated with Darby and Karni (1973) and initially focused on strategies to reduce fraud by credence-goods sellers (e.g., automobile mechanics, electronic specialists, and physicians). For example, Dranove (1988) studies the negative effect of unnecessary treatments on reputation and shows the extent of unnecessary treatment depends not only on the price and potential medical benefit of the treatment, but also on the relative diagnostic skills of the physician and the patient. Wolinsky (1993) shows reputation considerations can help discipline experts in these markets. Subsequently, Arora and Fosfuri (2005) study buyers' willingness to pay for diagnostic information that can be used for decision-making and suggest a nonmonotonic relationship between the buyer's willingness to pay and the ex-ante probability of success. Durbin and Iyer (2009) show the presence of side payments can facilitate truthful communication by a corruptible diagnostician. Jiang et al. (2014) show a diagnostic expert's pricing decision can function as a signal of the expert's altruism. Dai and Singh (2020) study how a physician can use the diagnostic process to signal his or her diagnostic skills to peers, and show undertesting can serve as a signal of diagnostic ability. Our paper contributes to this literature by studying the externality of individual diagnosis decisions, which has implications for both short- and long-term demand for diagnostic testing and poses a novel economic trade-off. In doing so, it enriches this literature by formalizing the interaction between diagnostic decision-making and testing capacity.

Another stream of related literature concerns information disclosure and emerges from the seminal work by Grossman (1981) and Milgrom (1981), who establish full unraveling of private quality information. Subsequent work has revealed a myriad of drivers of *partial* information disclosure. For example, Jovanovic (1982) and Verrecchia (1983) argue the presence of disclosure costs can result in partial disclosure, whereas Dye (1985) shows a manager who is believed to be imperfectly informed may be able to hide information. Bhardwaj et al. (2008) and Mayzlin and Shin (2011) reason a limited disclosure bandwidth may force a player to disclose only a subset of information. Guo and Zhao (2009) show competition can reduce firms' incentive to disclose information. Iyer and Singh (2018) find a

firm has a lower incentive to disclose product safety information using voluntary certification if product safety and consumer effort are complements but not when they are substitutes. Iyer and Singh (2022) show competing firms have lower incentives to disclose rival firms' negative information than to disclose their own positive information. Similar to our paper, Zhang (2014) studies implications of a mandatory-disclosure requirement imposed by the policymaker and shows mandatory disclosure does not always benefit consumers. However, the focus of that paper is on the policy's effect on consumers' inferences of quality.[5] In our paper, the lab conceals patient information to create additional future demand for testing. Because this information-concealment behavior also helps reduce the spread of disease transmission, mandatory disclosure may be undesirable even from the social planner's perspective.

Our research contributes to the growing theory literature on pandemic-related policy issues. Ely et al. (2021) examine a policymaker's decision to allocate tests that differ in their sensitivities and specificities and provide an algorithmic solution for the test-allocation problem. de Véricourt, Gurkan, and Wang (2021) explore governments' decision of communicating the severity of the pandemic and show the relative focus on the economy and public health can lead a government to downplay or exaggerate the severity of the pandemic. Guo (2022a) investigates whether and when a social planner's communication about the severity of the pandemic can be credible in the presence of misrepresentation incentives. Guo and Xu (2023) argue in support of more equal consumption of pandemic-fighting resources and suggest discriminatory subsidies in favor of individuals with lower willingness to pay for these resources can maximize social welfare. Guo (2022b) compares the effects of lockdowns and quarantines and shows that in some cases, a lockdown (or targeted isolation) may lead to worse consequences than quarantines. Our work is also related to several papers in the healthcare operations management literature related to (1) diagnostic testing policies (e.g., Adida and Dai 2024; Jain et al. 2024) and (2) pandemic control policies (e.g., Mak et al. 2022; Wang et al. 2024). We contribute to this literature by examining a policymaker decision of whether to mandate the reporting of the CT cutoff for PCR tests commonly used to diagnose infectious diseases.

[5]Another related stream of literature (e.g., Gal-Or et al. 2007; Guo 2020; Sun and Tyagi 2020) studies information-disclosure incentives in the context of distribution channels.

Finally, our paper also contributes to the literature on the delegation of decision rights. Lal (1986) and Bhardwaj (2001) examine the implications of the delegation of pricing decisions in the context of salesforce. Lal (1986) shows delegation can be desirable when the salesperson is better informed than the manager. Bhardwaj (2001) shows it can be desirable when price competition in the market is more intense. Singh (2017) studies how delegation of decision rights to an agent shapes corruptible behavior in a procurement auction. Özer et al. (2018) study the consequences of information sharing, advice provision, and delegation on cooperation between two parties. Dogan et al. (2018) show a firm's organization structure (centralized vs. decentralized) is an important driver of decision-making rights of its managers. In our paper, the lab decides whether to keep the right of deciding the patient's diagnosis or disclose the viral load on the test result and delegate the authority to physicians. We find the lab prefers to keep the decision right to itself.

3. Model

We consider a model with two periods (1 and 2) such that the same mass 1 of individuals live in both periods.[6] At the beginning of period 1, a proportion ϕ of these individuals are infectious (i.e., they can infect others with the disease on contact). The remaining $(1-\phi)$ are not infectious, either because they have never been infected or because they carry a negligible amount of viral residue, which can neither make them sick nor infect others.[7] After a patient becomes infectious, symptoms (e.g., fever, cough, headache, fatigue, sore throat, and new loss of taste or smell) may take several days to appear. In addition, some patients may fully recover without ever showing any symptoms (Buitrago-Garcia et al. 2020). We assume a proportion α of all infectious patients exhibit disease symptoms. The remaining proportion $(1-\alpha)$ of infectious patients are asymptomatic. The symptoms may not be unique to the disease; patients can experience them due to other diseases (e.g., influenza). A proportion β of $(1-\phi)$ individuals who are not infectious (where $\beta < \alpha$) also exhibit disease symptoms.

[6] Similar to Chen et al. (2022), we consider a two-period, discrete-time setting for tractability.

[7] In practice, some individuals (e.g., with a mass of λ) may have a low viral load in period 1 but become infectious during period 2. In this case, the mass of infections in period 2 will be increased by λ. Because such an increase does not change our main insights, we assume $\lambda = 0$ throughout the paper.

Patients with symptoms seek testing for the virus.

A lab performs a PCR test (hereafter, "test") to determine if a patient is infectious.[8] The test returns a CT value, denoted by x, a proxy for viral load. We assume x is continuous and normalize the range of x to $[0, 1]$.[9] The value of x returned by the test is stochastic and depends on whether the patient is infectious. A higher value of x indicates a lower viral load, and vice versa. For an infectious patient, the value of x is an independent draw from the triangular distribution with mode zero, probability density function (pdf) $g(x) = 2(1-x)$, and cumulative distribution function (cdf) $G(x) = 1-(1-x)^2$. Similarly, for a patient who is not infectious, the test returns an x that is an independent draw from a triangular distribution with mode one, pdf $h(x) = 2x$, and cdf $H(x) = x^2$. The CT value from a non-infectious patient's test result first-order stochastically dominates that from an infectious patient's test result; that is, the cumulative distribution function has a higher value for an infectious individual than for a non-infectious individual at any CT value x. Both conditional distributions have support of $[0, 1]$. The CT value (x) returned by the test is the private information of the lab.

The lab sets a CT cutoff $\kappa_l \in [0, 1]$ and concludes the patient is positive if $x \leq \kappa_l$ and negative if $x > \kappa_l$. Because a higher value of x indicates a lower viral load, the CT cutoff κ_l essentially sets an upper bound for a positive diagnosis. Most of the U.S. commercial labs (e.g., Quest Diagnostics) are for-profit entities (Foley 2020). However, because they are providing an essential public service (diagnostic medical testing), they are not necessarily entirely profit-driven. Consistent with the health economics literature (e.g., Arrow 1963), we assume the lab is motivated by both profitability and patient well-being in its decision-making. (We provide the expression for the lab's objective function in the next section.) The lab decides whether it will report only the binary diagnosis or both the diagnosis and CT value

[8]Most COVID-19 diagnostic labs (e.g., Labcorp and Quest) in the U.S. use their own PCR tests; our model is consistent with this practice. Our key findings carry over qualitatively to an extension in which the PCR test manufacturer and the lab are two separate players.

[9]Although a cycle number is, by definition, an integer, a patient's CT value can be fractional, because it measures the number of cycles at which the fluorescence passes the threshold, so it can be between two cycles. For example, according to a *New York Times* story (Grynbaum 2020) published on October 14, 2020, "Mr. Trump's P.C.R. test had a cycle threshold—a proxy for viral load—of 34.3, Dr. [Anthony] Fauci said." The CT value led Dr. Fauci to conclude, "We can say with a high degree of confidence that he is not transmissible."

on the test results.

Suppose the lab decides to include both the binary diagnosis and the CT value on the test results. In this case, the physician, who, unlike the lab, is only concerned with the well-being of the patient and understands the lab's decision may have been influenced by profit motives, disregards the lab's binary diagnosis and uses the CT value to diagnose the patient. The physician sets the CT cutoff $\kappa_p \in [0, 1]$ —the subscript p stands for "physician"—and diagnoses the patient as positive if $x \leq \kappa_p$ and negative if $x > \kappa_p$.[10] (We use "physician" as a proxy for individuals who act in the best interest of the patient. In real life, these individuals can include nurses, contact tracers, and caregivers.) An infectious patient may be diagnosed as negative (i.e., false negative) with probability $[1 - G(\kappa_p)]$. The infectious patient incurs a cost c_{FN} if she is diagnosed as negative; the cost includes, for example, the risk posed to the patient's family members and other close contacts. The patient's expected cost from a false-negative diagnosis is therefore $c_{FN} \cdot [1 - G(\kappa_p)]$. Similarly, a patient who is not infectious is diagnosed as positive with probability $H(\kappa_p)$. We represent the patient's cost in the event of a false-positive diagnosis by c_{FP}, which can include, for example, the patient's lost productivity and unnecessary isolation from family (Wu 2020). The patient's expected cost from a false-positive diagnosis is therefore $c_{FP} \cdot H(\kappa_p)$. True-positive and true-negative diagnoses do not impose any misdiagnosis costs.

Next, we describe the approach we follow to determine the CT cutoff. Several approaches have been proposed to set cutoffs for medical diagnostic testing, and they yield comparable cutoffs. The most common way to set cutoffs is based on the receiver operating characteristic (ROC) curve approach; see, for example, McNeil et al. (1975) for a seminal discussion of the approach and Ekelund (2011) for a more recent illustration. In its simplest form, the requirement can be described as follows: if a single test is being used for diagnosis, the cutoff should be set to produce equal test sensitivity ($s_e = G(\kappa_p)$: the probability that an infectious patient is tested positive) and specificity ($s_p = 1 - H(\kappa_p)$: the probability that a non-infectious patient is tested negative). Alternatively, the cutoff may be set to maximize the

[10] We define a true-positive condition as one in which the patient has a meaningful viral load such that further treatment and contact-tracing efforts are necessary. A true-negative condition does not necessarily mean the patient has zero viruses; it simply means the patient does not have enough viruses to be infectious or develop severe conditions.

sum $(s_e + s_p)$ of test sensitivity and specificity (or, equivalently, minimize the probability of erroneous diagnosis, that is $(1 - s_e) + (1 - s_p)$). Note the underlying assumption is that the costs of a false negative, c_{FN}, and false positive, c_{FP}, diagnosis are comparable. For a variety of diseases (including COVID-19), $c_{FP} \neq c_{FN}$. Therefore, we consider the cutoff as being set to minimize the sum of the expected cost of false-negative diagnosis and false-positive diagnosis, which we express as

$$\xi(\kappa_p) = (1 - s_e)c_{FN} + (1 - s_p)c_{FP}. \tag{1}$$

Substituting the expressions for s_e and s_p, we get

$$\xi(\kappa_p) = c_{FN} \cdot [1 - G(\kappa_p)] + c_{FP} \cdot H(\kappa_p). \tag{2}$$

Note, minimizing $\xi(\kappa_p)$ is equivalent to maximizing the patient utility under the assumption of maximum diagnostic uncertainty for all patients, because the benefits of true-positive and true-negative diagnosis can be absorbed into c_{FN} and c_{FP}, respectively. The physician, who acts in the best interest of the patient, chooses a CT cutoff of κ_p that minimizes $\xi(\kappa_p)$. Solving the first-order condition yields a test cutoff of

$$\kappa_p = \frac{c_{FN}}{c_{FN} + c_{FP}}. \tag{3}$$

Note the above approach to determine test cutoffs is agnostic about the extent of disease prevalence in the population, a property that has been acknowledged as a significant benefit of the ROC approach, particularly in scenarios of rapidly evolving disease spread (Doi 2013). This approach assumes the maximum diagnostic uncertainty, that is, the sample being tested is equally likely to be positive or negative. Alternative approaches in which the cutoffs respond to the extent of disease spread have been proposed, but they face many implementation challenges (Bentley et al. 2012; Caraguel et al. 2011). For example, if cutoffs were to be a function of the COVID-19 spread level in the population, different geographic and demographic groups would be diagnosed using different cutoffs, which would need to change with time as different waves of infection hit the population. Doing so, however, would

create significant implementation issues and confusion. Therefore, consistent with medical practice in the US, in our main analysis, we follow an approach in which the test cutoff is non-responsive to the level of virus spread. We are not aware of any practice in which testing labs dynamically update their diagnostic threshold based on the disease prevalence among the population. Our main results are robust to alternative approaches for setting test cutoffs; see Section 5.1 for a robustness check for the case in which the CT cutoff depends on the disease prevalence. All patients whom the physician classifies as positive are quarantined.

Now, suppose the lab decides to report only its binary diagnosis on the test result. In this case, the physician does not receive the CT-value (x) information, and patients are quarantined following the lab's binary diagnosis. By not disclosing the CT value on the test result, the lab essentially keeps the diagnostic-decision right with itself. By contrast, by disclosing the CT value, the lab forfeits its diagnostic-decision right to the physician. A lab's unwillingness to disclose the CT value does not necessarily mean the CT value will not become available to the physician. A policymaker decides whether to require the lab to disclose the CT value to physicians.[11] If the policymaker mandates CT-value reporting, the lab follows the policymaker's directions and discloses it on test results.

In addition to deciding the CT-value disclosure policy and CT cutoff (κ_l) in period 1, the lab decides its testing capacity C_2 for period 2 (which is also the terminal period); we assume $C_2 \geq C_1$ and the special case of $C_2 = C_1$ means the lab does not invest in building additional capacity. Building testing capacity is costly and time consuming because "buying new machinery is expensive, and staff have to be retrained and incentivized to work in tough conditions"; it is also risky due to demand uncertainty, because "it's hardly appealing to build capacity [the lab] might not need six months from now" (Foley 2020). The cost of period-2 testing capacity is given by $\gamma(C_2^2 - C_1^2)$,[12] where $\gamma > 0$ is the capacity-cost parameter.[13] (We normalize the unit cost of the test kit to zero for simplicity of analysis.) After setting its

[11]The policymaker has no direct control over the testing process or the interpretation of testing results, which is considered the practice of medicine in the U.S. and thus cannot be explicitly regulated (Shirts 2020).

[12]We have formally examined an alternative model in which the cost of developing a period-2 testing capacity of C_2 is specified as $\gamma(C_2 - C_1)^2$ and we find our main results continue to hold.

[13]A convex cost structure is aligned with the scenario of building laboratory testing capacity, which often entails purchasing new machines and hiring new personnel. We also note the literature commonly assumes a convex cost function (e.g., a quadratic function) for capacity; see a discussion by Eliashberg and Steinberg (1993, §3).

viral-load disclosure policy, CT cutoff, and period-2 capacity, the lab starts testing patient samples.

For each tested sample, the lab generates a revenue of r_0. Because r_0 is usually determined by the insurance company and is independent of the lab's disclosure decision or the volume of tests processed, we assume it is exogenously given. In period 1, the lab has a limited and exogenously given testing capacity C_1, which is insufficient to meet the overwhelming period-1 demand for testing (specifically, $C_1 < \alpha\phi + (1-\phi)\beta$). Therefore, a subset of patients is randomly selected from those seeking testing (i.e., from those showing the disease symptoms in period 1) to match the testing capacity.

All patients with positive test results are quarantined; these patients include both true and false positives. Once quarantined, true-positive patients do not infect others, and false-positive patients do not become infected. True-negative patients (i.e., those who carry no or negligible viral load) are not infectious.[14] However, certain infectious patients are either not tested or tested false-negative in period 1. Because these patients are not quarantined, they come into contact with others and spread the virus. We assume each such patient infects R_0 additional individuals. (In epidemiology, R_0 is known as the basic reproduction number and captures the infectiousness of the disease.) We assume R_0 follows a uniform distribution with a support $[0, \bar{R}_0]$ and cdf $F(R_0) = R_0/\bar{R}_0$. We assume all the ϕ patients, who are infectious in period 1, recover at the end of period 1. The rationale behind this assumption is the length of time (up to several months) needed for testing-capacity changes compared with the weeks needed for patients to fully recover from the infection.[15]

In period 2, in addition to those with symptoms, individuals who came into direct contact with those who tested positive in period 1 seek to be tested. We refer to the demand from this group as *contact-tracing demand* (or *diagnosis-induced demand*). It is determined by the contact-tracing intensity ρ, which represents the number of individuals who seek testing because they came into direct contact with someone who subsequently tested positive in period 1.[16] Similar to period 1, if demand for testing is higher than the period-2 testing

[14] As highlighted by Dr. Anthony Fauci in his interview with *This Week in Virology,* culturing viruses is almost impossible if the CT value is higher than 35 (TWiV 2020).

[15] A formal examination of an alternative model specification in which period-1 patients continue to remain sick in period 2 reveals our main insights continue to hold.

[16] Note ρ measures demand for additional testing as an immediate result of a positive diagnosis. Such

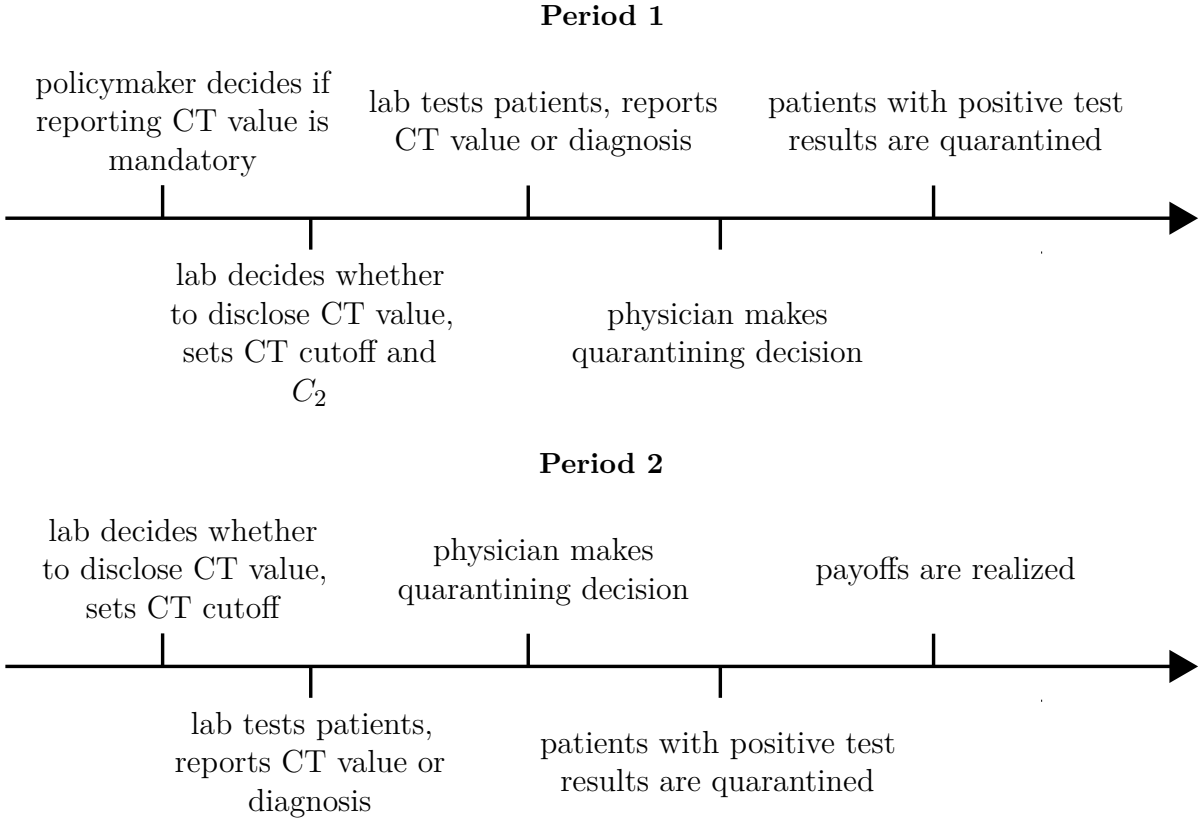


Figure 1: Timing

capacity, a subset of all individuals seeking testing is randomly selected to match the testing capacity C_2. The lab has a discount factor of 1.

The timing of actions (also shown in Figure 1) is as follows: all the policy and capacity decisions are made in period 1, which has five stages. In the first stage, the policymaker decides whether the disclosure of the CT value is mandatory. Next, the lab decides whether to disclose the CT value and sets the CT cutoff and the period-2 capacity. Third, period-1 patients are tested, and the CT value or diagnosis is shared with the physician. Fourth, if the lab discloses the CT value from the test, the physician determines (based on the physician's own CT cutoff) whether the patient is positive, whereas if the lab only shares the diagnosis and not the CT value, the physician follows the lab's diagnosis. Fifth, patients who test positive are quarantined. At the beginning of period 2, the demand for testing is realized

demand can come from formal contact-tracing efforts as well as from individuals' voluntary testing decisions. For the rest of the paper, for ease of exposition, we use "contact tracing" to refer to all such efforts that contribute to diagnosis-induced demand.

ϕ	the proportion of individuals who are infectious at the beginning of period 1
ϕ_2	the mass of infectious patients at the beginning of period 2
α	the proportion of infectious patients who exhibit disease symptoms
β	the proportion of non-infectious individuals who exhibit disease symptoms
x	the CT value ($x \in [0,1]$)
$G(x)$	the cumulative distribution function of the value of x for an infectious patient
$H(x)$	the cumulative distribution function of the value of x for a non-infectious patient
κ_l	the lab's CT cutoff
κ_p	the physician's CT cutoff
c_{FP}	the patient's cost in the event of a false-positive diagnosis
c_{FN}	the patient's cost in the event of a false-negative diagnosis
r_0	the lab's revenue from each tested sample
C_1	the lab's testing capacity in period 1
C_2	the lab's testing capacity in period 2
R_0	the infectiousness of the disease, which is uniformly distributed
$\bar{R}_0$	the upper bound of the range of R_0
ρ	the intensity of contact tracing
D_2	the demand for testing in period 2
s_2	the mass of patients with symptoms in period 2
$\xi(\kappa)$	the patient's expected costs from false-positive and false-negative diagnoses under a CT cutoff of κ
$\pi(\kappa, C_2)$	the lab's objective function under a CT cutoff of κ and a period-2 capacity of C_2

Table 1: Notation

based on contact-tracing intensity and the number of patients with symptoms. The lab makes CT-value-disclosure and CT-cutoff decisions. The next two stages of period 2 are similar to the third and fourth stages of period 1. Finally, patients who test positive are quarantined, and payoffs are realized. All players are risk neutral. We solve this sequential-move game using backward induction.[17]

At this point, we note that our main model considers a specific setting (with a for-profit lab and the physician caring only about patient welfare) for simplicity. In the real world, some labs (e.g., government and university labs) may be non-profit organizations that care

[17]Consistent with the practice of medicine in the US, when setting the CT cutoff both the lab and the physician consider the tested patient is equally likely to be positive or negative. However, this assumption can be seen as a deviation from standard economic theory in the sense that players with knowledge (or rational expectation) of the disease-prevalence level ignore it while setting the CT cutoff. In addition, when the lab does not disclose the CT value, the physician uses the period-1 CT cutoff to compare the patient's expected cost from following (assuming maximum diagnostic uncertainty) and dismissing the lab's diagnosis. See the model extension presented in Section 5.1 which explicitly considers disease prevalence when calculating CT cutoffs.

about patients and the society, but not revenues. In addition, different stakeholders (e.g., physician, care givers, contact tracers, patients, etc.) may differ in the extent to which they care about the patient and the society. In Section 5.2, we develop an alternative model of overdiagnosis (assuming away undertesting), in which the lab is not concerned about its own profitability. We capture the idea that different stakeholders may put different weights on the costs imposed on the patient and the society. Our analysis suggests the robustness of our overdiagnosis result. We summarize the notation that we use in Table 1.

4. Analysis

In this section, we analyze our model and generate managerial and policy insights. To begin with, in Section 4.1, we characterize the lab's decisions with regard to its optimal diagnostic threshold and testing capacity. In Section 4.2, we show the results provide a theory that explains the co-existence of overdiagnosis and undertesting. Finally, in Section 4.3, we analyze the social planner's problem of whether to require the lab to disclose the CT value on their test results.

4.1 Diagnostic Threshold and Testing Capacity

First, we characterize the demand for testing in period 2 (i.e., D_2) for a given CT cutoff (i.e., κ_l) set by the lab in period 1 under non-disclosure of the CT value. We denote by s_1 the mass of patients with symptoms in period 1, which is given by $\phi\alpha + (1-\phi)\beta$; all these patients seek testing. However, because initial testing capacity is limited (specifically, $C_1 < s_1$), only a subset of these patients are tested. A total of C_1 patients are randomly selected and tested. As a result, among the patients who are tested in period 1, the probability that a patient is infectious is $\frac{\alpha\phi}{\alpha\phi+(1-\phi)\beta}$.

For a given CT cutoff κ_l, the probability that the lab classifies a patient as positive (i.e., infectious) is given by

$$\Pr(x \leq \kappa_l) = \frac{\alpha\phi}{\alpha\phi + (1-\phi)\beta} G(\kappa_l) + \left[1 - \frac{\alpha\phi}{\alpha\phi + (1-\phi)\beta}\right] H(\kappa_l) \tag{4}$$

$$= \frac{\alpha\phi\left(2\kappa_l - \kappa_l^2\right) + (1-\phi)\beta\kappa_l^2}{\alpha\phi + (1-\phi)\beta}, \tag{5}$$

and the probability that the lab classifies a patient as negative is

$$\Pr(x > \kappa_l) = \frac{\alpha\phi}{\alpha\phi + (1-\phi)\beta}\left[1 - G(\kappa_l)\right] + \left[1 - \frac{\alpha\phi}{\alpha\phi + (1-\phi)\beta}\right]\left[1 - H(\kappa_l)\right] \tag{6}$$

$$= \frac{\alpha\phi(1-\kappa_l)^2 + (1-\phi)\beta\left(1-\kappa_l^2\right)}{\alpha\phi + (1-\phi)\beta}. \tag{7}$$

Among the patients who are tested in period 1, the probabilities that a patient receives a false-positive diagnosis and a false-negative diagnosis, as functions of κ_l, are $\frac{(1-\phi)\beta}{\alpha\phi+(1-\phi)\beta}\kappa_l^2$ and $\frac{\alpha\phi}{\alpha\phi+(1-\phi)\beta}(1-\kappa_l)^2$, respectively. As expected, a higher κ_l reduces the probability of a false-negative diagnosis but increases the probability of a false-positive diagnosis.

Among the C_1 patients who are tested in period 1, a fraction, $\frac{\alpha\phi}{\alpha\phi+(1-\phi)\beta}\left(2\kappa_l - \kappa_l^2\right)$, test positive and are actually infectious (true positives). Recall that we normalize the population size to one, so a total of ϕ individuals are infectious at the start of period 1; of them, only those who test positive are quarantined, so a mass $\phi - \frac{\alpha\phi}{\alpha\phi+(1-\phi)\beta}\left(2\kappa_l - \kappa_l^2\right)C_1$ of patients will continue to spread the virus. Thus, the mass of infectious patients at the beginning of period 2 is

$$\phi_2 = \left[\phi - \frac{\alpha\phi}{\alpha\phi + (1-\phi)\beta}\left(2\kappa_l - \kappa_l^2\right)C_1\right]R_0. \tag{8}$$

A fraction α of these infectious patients are symptomatic. In addition, a fraction β of $(1-\phi_2)$ individuals (who are not infectious with the virus) also show symptoms. Therefore, the mass of patients with symptoms in period 2 is

$$s_2 = \phi_2\alpha + (1-\phi_2)\beta. \tag{9}$$

All s_2 patients seek testing. Additional demand for testing is generated in period 2 as a result of contact-tracing efforts. Close contacts (e.g., family members, friends, and colleagues) of patients who tested positive in period 1 also seek testing. Therefore, the total demand for

testing in period 2 is

$$D_2 = s_2 + \rho C_1 \cdot \frac{\alpha\phi\left(2\kappa_l - \kappa_l^2\right) + (1-\phi)\,\beta\kappa_l^2}{\alpha\phi + (1-\phi)\,\beta}, \tag{10}$$

where the fraction is the probability that each of the C_1 patients undergoing testing receives a positive diagnosis in period 1, and ρ is the intensity of contact tracing. (In reality, additional demand for testing may arise for vulnerable groups, such as healthcare workers, teachers, and other essential workers, who may be required to undergo routine testing. Such additional demand can be captured by adding a constant term to (10) and does not change any of our key insights.) We can rearrange (10) to separate the effects of contact tracing and the contagious nature of the infection as

$$D_2 = z_1(\kappa_l) + z_2(\kappa_l) R_0, \tag{11}$$

where

$$z_1(\kappa_l) \triangleq \beta + \rho C_1 \cdot \frac{\alpha\phi\left(2\kappa_l - \kappa_l^2\right) + (1-\phi)\,\beta\kappa_l^2}{\alpha\phi + (1-\phi)\,\beta}, \tag{12}$$

$$z_2(\kappa_l) \triangleq (\alpha - \beta)\left[\phi - \frac{\alpha\phi}{\alpha\phi + (1-\phi)\,\beta}\left(2\kappa_l - \kappa_l^2\right) C_1\right]. \tag{13}$$

Note $z_1(\kappa_l)$ also includes the constant testing demand of β from non-infectious patients who show symptoms, in addition to diagnosis-induced demand.

Lemma 1. *$z_1(\kappa_l)$ increases in κ_l, whereas $z_2(\kappa_l)$ decreases in κ_l.*

Appendix A1 contains the proof of Lemma 1, along with the proofs of other technical results. Lemma 1 suggests a higher CT cutoff κ_l set by the lab increases the mass of patients who test positive. As a consequence, diagnosis-induced demand for testing increases. Additionally, a higher CT cutoff κ_l facilitates quarantining of a larger share of infectious patients (because true positives also increase), thus dampening the spread of the virus and putting a downward force on the demand for testing in period 2.

In formulating the lab's objective function, we omit its revenue from the first period because it is fixed. In period 2 (which is also the terminal period), the lab is not concerned

about its future demand. Therefore, the incentives of the lab and the physician are perfectly aligned in period 2. The lab sets a CT cutoff of $c_{FN}/\left(c_{FN}+c_{FP}\right)$ that minimizes the sum of the patient's expected false-positive and false-negative costs (i.e., $c_{FP}\cdot H(\kappa_l)+c_{FN}\cdot[1-G(\kappa_l)]$). Equivalently, in period 2, the lab reports the CT value on the test result. Note both the CT-cutoff value and the patient's expected cost of erroneous diagnosis from the second period are independent of the lab's period-1 decisions. For ease of notation, we define

$$r \triangleq r_0 - \min_{\kappa_l}\{c_{FP}\cdot H(\kappa_l)+c_{FN}\cdot[1-G(\kappa_l)]\}. \tag{14}$$

Thus, in period 1, the lab sets κ_l and C_2 to maximize its objective function:[18]

$$\begin{aligned}\pi(\kappa_l, C_2) = \; & r\cdot \mathbb{E}\min\{C_2, z_1(\kappa_l)+z_2(\kappa_l)\cdot R_0\} \\ & -\gamma\cdot(C_2^2-C_1^2)-C_1\cdot\{c_{FP}\cdot H(\kappa_l)+c_{FN}\cdot[1-G(\kappa_l)]\},\end{aligned} \tag{15}$$

where the first term represents the lab's total expected revenue less patients' expected cost of erroneous diagnosis during period 2, the second term is the cost of building capacity, and the third term captures patients' expected cost of erroneous diagnosis during period 1.[19]

At this point, understanding how the lab's incentive to set the period-2 capacity C_2 depends on the period-1 CT cutoff κ_l is useful. The following proposition describes the lab's optimal capacity choice for an exogenously given κ_l.

Proposition 1. *If the lab follows an exogenous κ_l, its optimal capacity choice $C_2^*(\kappa_l)$ is*

$$C_2^*(\kappa_l) = \frac{r[z_1(\kappa_l)+z_2(\kappa_l)\bar{R}_0]}{r+2z_2(\kappa_l)\bar{R}_0\gamma}, \tag{16}$$

which increases in κ_l if

$$\rho \geq \frac{\phi\alpha\left(\alpha-\beta\right)\left(1-\kappa_l\right)\bar{R}_0}{\phi\alpha\left(1-\kappa_l\right)+\left(1-\phi\right)\beta\kappa_l}. \tag{17}$$

[18] To simplify the analysis, we assume the lab weighs each dollar of welfare loss to patients equal to each dollar earned from testing. In an alternative model presented in Section 5.2, we consider the case of a non-profit lab that does not care about its own profitability and show that overdiagnosis continues to exist.

[19] We focus on the case in which γ is not prohibitively high such that $C_2 \geq z_1(\kappa_l)$ in equilibrium. Otherwise, the term $\min\{C_2, z_1(\kappa_l)+z_2(\kappa_l)\cdot R_0\}$ in (15) will always be equal to C_2 and the lab's problem will become a trivial one.

If the intensity of contact tracing is sufficiently large, the lab's period-2 capacity increases in the CT cutoff. A larger CT cutoff leads to more patients testing positive (including false positives). In the presence of a significant contact-tracing effort, a large number of positive diagnoses leads to a high diagnosis-driven demand for testing in period 2 and incentivizes the lab to invest more in period-2 capacity. The effect of a larger expected R_0 is the opposite. If the virus is highly contagious, a high demand for period-2 testing is organically generated. In this case, a higher CT cutoff, which increases the true-positive rate, imposes a large cost on the lab in the sense that a larger proportion of infectious patients are quarantined and stop contributing toward increasing period-2 demand for testing (i.e., they stop spreading the virus). However, if the virus is not highly contagious (expected R_0 is small), the lab becomes more motivated to set a higher CT cutoff, which boosts contact-tracing demand but does not significantly hurt the organic demand for testing in period 2. Therefore, the lab becomes more motivated to invest in building period-2 capacity.

Proposition 1 relies on a sufficient (but not necessary) condition that the intensity of contact tracing is sufficiently large. If this sufficient condition is not satisfied, the lab's optimal capacity choice may either increase or decrease in κ_l.

Suppose the policymaker requires the lab to report the CT value in the test result. The following lemma describes the physician's CT cutoff and the lab's period-2 capacity decisions in this scenario.

Lemma 2. *If the lab is mandated to report the CT value on the test result, the physician diagnoses the patient using a CT cutoff* $\kappa_p = c_{FN}/\left(c_{FN}+c_{FP}\right)$ *and the lab sets its period 2 capacity at*

$$C_2^*\left(\kappa_p\right) = \frac{r[z_1(\kappa_p)+z_2(\kappa_p)\bar{R}_0]}{r+2z_2(\kappa_p)\bar{R}_0\gamma}. \tag{18}$$

If the CT-value information is available to the physician, she uses this information to diagnose the patient using her own CT cutoff κ_p, which minimizes the sum of the patient's expected false-positive and false-negative costs. Because the physician's objective function and costs are common knowledge, both the lab and the policymaker can anticipate this κ_p. The lab sets its period-2 capacity at $C_2^*\left(\kappa_p\right)$, as given by equation (18).

Next, suppose the policymaker does not require the lab to report the CT value on the

test result. In this case, the lab decides whether to disclose the CT value voluntarily or to disclose only the diagnosis based on its own CT cutoff. If the lab chooses to voluntarily report the CT value on the test result, similar to the case in which the lab is mandated to report the CT value, the physician sets the CT cutoff κ_p at $c_{FN}/\left(c_{FN}+c_{FP}\right)$ and the lab sets period-2 testing capacity at $C_2^*\left(\kappa_p\right)$. The following proposition characterizes the lab's optimal decisions (κ_l^* and C_2^*) when the lab sets its own CT cutoff and reports only the test outcome (positive or negative).

Proposition 2. *If the lab sets its own CT cutoff and the physician follows the lab's diagnosis, the lab's optimal capacity C_2^* is*

$$C_2^* = \frac{[z_1(\kappa_l^*) + z_2(\kappa_l^*)\bar{R}_0]r}{r + 2z_2(\kappa_l^*)\bar{R}_0\gamma}, \tag{19}$$

where κ_l^* satisfies

$$\frac{r}{2\bar{R}_0 C_1}\int_0^{\frac{\bar{R}_0[r-2\gamma z_1(\kappa_l^*)]}{r+2\gamma\bar{R}_0 z_2(\kappa_l^*)}} [z_1'(\kappa_l^*) + z_2'(\kappa_l^*)R_0]dR_0 = c_{FP}\kappa_l^* - c_{FN}(1-\kappa_l^*). \tag{20}$$

The above solution to the lab's problem (κ_l^* and C_2^*) has implications for the lab's tradeoff in choosing its CT cutoff and testing capacity. Note from (20) that if $\kappa_l^* = \kappa_p = c_{FN}/(c_{FN}+c_{FP})$, the right-hand side of the equation will be equal to zero. However, because the term $z_1'(\kappa_l^*) + z_2'(\kappa_l^*)R_0$ in the left-hand side of (20) is not necessarily equal to zero, the lab may not set its CT cutoff κ_l^* the same as κ_p. Note, the lab's CT cutoff κ_l^* is decreasing in γ and increasing in r, ρ, and C_1. (See the proof of Proposition 4 (ii) in the Appendix.) The reason is that a higher cost of building capacity (γ) reduces the lab's incentive to boost its period-2 demand. Therefore, the lab sets a smaller CT cutoff. Similarly, a higher revenue (r), a higher contact-tracing rate (ρ), or a higher period-1 capacity (C_1) increases the lab's incentive to boost its period-2 demand. In this case, the lab raises its CT cutoff. In the following lemma, we specify the condition that ensures the physician follows the lab's diagnosis.

Lemma 3. *The physician follows the lab's diagnosis if*

$$c_{FN}\left(1-\kappa_l^*\right)^2 + c_{FP}(\kappa_l^*)^2 \leq \min\{c_{FN}, c_{FP}\}. \tag{21}$$

Even though the physician does not know the CT value, she knows its distribution and can thus use the CT cutoff to calculate the expected cost of following versus ignoring the lab's diagnosis to each patient. She follows the lab's diagnosis if doing so results in a lower expected cost of erroneous diagnosis for the patient. The lab understands the physician may reject its diagnosis if κ_l^* is too extreme. In the equilibrium, the lab's κ_l^* is such that the physician follows the lab's diagnosis. As shown in the proof of Proposition 3, if $c_{FN} \geq c_{FP}$, which is often the case for infectious diseases, the condition specified in Lemma 3 is redundant. However, if $c_{FN} < c_{FP}$, it imposes an upper bound on κ_l^*. We compare κ_l^* and κ_p in the following subsection and draw implications for the lab's capacity decision.

4.2 Overdiagnosis and Undertesting

Having analyzed the lab's optimal diagnostic threshold and capacity, we are now ready to generate implications with regard to how the lab's decisions compare to the optimal decision from the patient's perspective. We also discuss whether the lab has an incentive to report the CT value alongside its binary diagnosis.

The following proposition presents a comparison of the CT cutoffs κ_l^* (which is optimal from the lab's perspective) and κ_p (which is ideal from the patient's perspective). It also compares the lab's period-2 testing-capacity decisions corresponding to the CT-cutoff values of κ_l^* and κ_p. Finally, it describes the lab's decision on whether to report the CT value in the absence of a mandatory reporting requirement.

Proposition 3. *If $\rho \geq (\alpha-\beta)\bar{R}_0$, in the absence of a mandatory CT-value reporting requirement, in period 1,*

(i) if the lab sets the optimal CT cutoff, it is higher than κ_p;

(ii) if the lab sets the optimal CT cutoff, the corresponding period-2 capacity is higher than the capacity under a CT cutoff of κ_p; and

(iii) the lab prefers to set the CT cutoff itself and diagnose the patient rather than report the CT value on the test result.

Before providing intuition behind Proposition 3, we remark about the condition $\rho \geq (\alpha - \beta)\bar{R}_0$. Note from footnote 16 that ρ captures the demand for additional testing resulting from a positive diagnosis; the demand can come from formal contact-tracing efforts, but more importantly, from voluntary testing based on an individual's assessment of his or her contact history, even in the absence of formal contact tracing. The average U.S. household size is 2.5 (US Census 2019), and it has been estimated that an infectious individual comes into contact with an average of three persons outside his or her home (Feehan and Mahmud 2021); accordingly, we estimate ρ at $2.5 \times (3+1) - 1 = 9$. In the particular case of COVID-19, the Centers for Disease Control and Prevention (CDC) estimated R_0 as 2.5, with an upper bound of 4.0, and the percent of infections that are symptomatic (α) as 60% (CDC 2020). Thus, we conclude the condition $\rho > \alpha\bar{R}_0 > (\alpha - \beta)\bar{R}_0$ can be satisfied under fairly reasonable parametric assumptions.

In terminal period 2, the CT cutoff is set at κ_p regardless of whether the lab or the physician sets it. The reason is that in period 2, the lab is not concerned about generating future demand for testing and, similar to the physician, acts in the best interest of the patient. However, in period 1, the lab sets a higher CT cutoff than what physicians would consider appropriate (i.e., $\kappa_l^* > \kappa_p$). In the proof of Proposition 3 (i), we show that at $\kappa = \kappa_l^*$ and in the absence of the CT-value information, the physician follows the lab's diagnosis to recommend isolation to the patient. A high CT cutoff helps the lab boost diagnosis-induced demand while suppressing demand from community spread. If the lab expects a sufficiently large number of close contacts of patients who tested positive to seek testing, it focuses on future demand creation by increasing the CT cutoff and thus engaging in overdiagnosis. This result and its intuition provide a rationale for the observed high CT cutoff that is currently set by labs and challenged by healthcare experts (Mandavilli 2020a, 2020b; Rao 2020; Service 2020). It also suggests that in the future, we can expect labs to either report the CT value on the test result or lower the CT-cutoff value when determining whether a patient tests positive.

The lab sets a higher period-2 testing capacity when setting its own period-1 CT cutoff

at κ_l^* than the capacity chosen corresponding to the CT cutoff of κ_p. This result builds on part (i) of the proposition: $\kappa_l^* > \kappa_p$. The lab sets a higher CT cutoff to boost its future demand for testing. Yet, for demand to translate into revenue, it must invest in building testing capacity. Put differently, an inflated CT cutoff is not entirely a negative force, because it incentivizes the lab to expand its testing capacity.

Next, we examine patient-welfare implications of the lab's decision to diagnose patients using its own CT cutoff κ_l^*. Recall from the Model section, the value of true-positive and true-negative diagnosis can be absorbed into c_{FN} and c_{FP}, respectively. Therefore, minimizing $\xi(\kappa)$, which is the sum of the patient's expected costs from false-positive and false-negative diagnosis, is equivalent to maximizing patient welfare. In period 2, the CT cutoff is set at κ_p regardless of whether the lab or the physician set it. Therefore, the expected welfare of the patient tested in period 2 is constant. The following proposition compares patient welfare corresponding to the period-1 CT cutoffs of κ_l^* and κ_p. In addition, the proposition presents how patient welfare (corresponding to the lab's period-1 CT cutoff κ_l^*) responds to changes in key model parameters γ, r, ρ, and C_1.

Proposition 4. *If* $\rho \geq (\alpha - \beta)\bar{R}_0$,

(i) $\xi(\kappa_l^*) > \xi(\kappa_p)$ *(equivalently, patient welfare is lower when the diagnosis is based on the lab's CT cutoff* κ_l^**) in period 1.*

(ii) Patient welfare (corresponding to the lab's period-1 CT cutoff κ_l^**) is increasing in* γ *and decreasing in* r*,* ρ*, and* C_1*.*

The patient welfare is maximized when the CT cutoff is set at $\kappa_p = c_{FN}/(c_{FN} + c_{FP})$. A CT cutoff that is higher than κ_p results in too many patients testing positive (a false-positive rate that is higher than what is optimal from the patient's perspective), which drives the patient welfare lower. As established in Proposition 3(i), in period 1, the lab sets $\kappa_l^* > \kappa_p$. Therefore, patient welfare is lower when the patient's diagnosis is based on the lab's CT cutoff κ_l^* than the physician's cutoff κ_p.

An increase in the lab's cost γ of adding capacity decreases the lab's incentive to boost its period-2 demand, by inflating the CT cutoff. The lab sets a lower κ_l^* in response to a

higher γ. Note this lower κ_l^* is still higher than κ_p. Because a decrease in κ_l^* reduces $\xi(\kappa_l^*)$, if $\kappa_l^* > \kappa_p$, the patient's expected cost is lower. Therefore, the patient welfare increases with γ. The main idea behind the above intuition is the following: if a change in some parameter reduces (increases) the lab's incentive to boost its period-2 demand, inflating the CT cutoff will result in higher (lower) welfare of the tested patient. An increase in (1) the net revenue from the test (r), (2) contact-tracing rate (ρ), or (3) period-1 capacity (C_1) incentivizes the lab to further inflate the CT value and results in a lower patient surplus.

4.3 Should CT-Value Reporting Be Mandatory?

If the policymaker does not require mandatory CT-value reporting, the lab is indifferent between reporting and not reporting the CT value as part of the test result in the terminal period 2. However, Proposition 3 shows that, in period 1, motivated by its desire to enhance period-2 demand, the lab finds *not* reporting the CT value is optimal, because the lack of a mandate allows the lab to inflate the CT cutoff (to a level above κ_p) when determining whether period-1 patients test positive.

Given the lab's decision to inflate the period-1 CT cutoff, healthcare experts' outcry against the common practice of not reporting CT values on test results begs the critical question of why policymakers do not intervene and require the lab to disclose CT values. The following proposition examines the implications of a mandatory CT-value reporting requirement and offers a possible rationale behind the policymaker's lack of intervention despite concerns raised by healthcare experts.

Proposition 5. *If $\rho \geq (\alpha - \beta)\bar{R}_0$, a mandatory CT-value reporting requirement imposed on the lab for period 1 results in*

(i) a higher expected utility of the tested patient,

(ii) a more widespread infection in the subsequent period, and

(iii) a lower period-2 testing capacity yet a higher service level (which is the likelihood that no testing shortage would occur)

than in the absence of a mandatory-reporting requirement.

If the reporting of CT value—a proxy for viral load—is not mandatory, the lab does not report it in period 1 and inflates the CT cutoff above what is best from an individual patient's perspective. A requirement to report the CT value as part of a test result in essence transfers the diagnostic-decision right from the lab to the physician. The physician, who acts in the best interest of the patient, sets the CT cutoff at κ_p. Therefore, as a result of mandatory CT value reporting requirement, the expected benefit to the patient tested increases.

One drawback of a mandatory viral-load reporting requirement in period 1 is that it leads to more false-negative diagnoses, potentially contributing to more widespread infection in the subsequent period. Although a higher CT cutoff imposes a cost on the tested patients, it facilitates quarantining of a larger proportion of infectious patients, thereby reducing community spread. To the extent society is willing to bear the costs associated with quarantining a large number of individuals who may not be contagious, in order to reduce community spread, letting the lab set a higher CT cutoff may be better than requiring viral-load reporting.

Finally, if the lab is not mandated to report the viral load in period 1, it sets the CT cutoff at κ_l^* and the period-2 capacity at C_2^*, which is higher than $C_2^*(\kappa_p)$. (See the related Proposition 3 *(ii)* and its intuition.) Although $C_2^* > C_2^*(\kappa_p)$, it is insufficient to absorb the increase in demand resulting from the lab's inflated CT cutoff κ_l^*. As the lab increases its CT cutoff, it finds delivering the same service level (by sufficiently increasing its capacity) is increasingly costly. An insufficient increase in the period-2 capacity essentially increases the likelihood of continued testing shortages in period 2. Interestingly, overdiagnosis not only coexists with undertesting, but also contributes to its increased likelihood.

5. Extension and Robustness of Results

In this section, we present results from two different models: The first model considers a disease prevalence dependent CT cutoff. The second model presents an alternative theory of overdiagnosis for a nonprofit lab with no capacity constraint.

5.1 Prevalence-Dependent CT Cutoff

In our main analysis, consistent with industry practice for setting cutoffs for medical diagnostic tests, we use the ROC approach to set the CT cutoff that minimizes the sum of the expected costs to the patient of false-positive and false-negative diagnoses. The implication is that the diagnostic criterion for medical tests is prevalence independent. In this section, for theoretical interest, we consider disease prevalence (i.e., the prior probability that a tested patient is positive) in determining the CT cutoff.

In period 1, the physician and the lab use their knowledge of the model parameters to calculate the probability that a tested patient is infectious as $\frac{\alpha\phi}{\alpha\phi+\beta(1-\phi)}$. The patient is not infectious with the complementary probability. Therefore, the physician's goal is to set a CT cutoff (κ_{p1}) that minimizes the patient's expected cost of an erroneous diagnosis:

$$\xi(\kappa_{p1}) = \frac{\alpha\phi}{\alpha\phi + \beta(1-\phi)} c_{FN}[1 - G(\kappa_{p1})] + \frac{\beta(1-\phi)}{\alpha\phi + \beta(1-\phi)} c_{FP} H(\kappa_{p1}). \tag{22}$$

Equivalently, the physician maximizes the patient's expected utility, since the benefits of a true-positive and true-negative diagnosis can be absorbed in c_{FN} and c_{FP}, respectively. Solving the first-order condition with respect to κ_{p1} gives the physician's period-1 CT cutoff:

$$\kappa_{p1} = \frac{\alpha\phi c_{FN}}{\alpha\phi c_{FN} + \beta(1-\phi)c_{FP}}. \tag{23}$$

As expected, the CT cutoff now depends on the parameters α, β, and ϕ. In addition, if $\phi = 0$, the cutoff is $\kappa_{p1} = 0$. Similarly, if $\alpha = 0$, the cutoff is $\kappa_{p1} = 0$.

In period 2, which is also the terminal period, the lab has no incentive to increase future demand. Therefore, the lab and the physician solve the same optimization problem to set the CT cutoff. If the lab does not disclose the CT value in period 1, the physician also incorporates information about the period-1 CT cutoff (which can be directly observed or rationally inferred) to obtain period-2 disease prevalence. In addition, when the lab does not disclose the CT value, the physician uses the period-1 CT cutoff to compare the expected cost of following and dismissing the lab's diagnosis to the patient. The lab sets the CT cutoff to ensure the physician does not dismiss the lab's test results. Note that, as specified in

(8), the mass of the infectious individual at the beginning of period 2 is ϕ_2. In addition, as specified in (10), the total demand for testing in period 2 is D_2. Therefore, the probability that a patient (tested in period 2) is infectious is $\frac{\alpha\phi_2}{D_2}$. The patient is not infectious with probability $1-\frac{\alpha\phi_2}{D_2}$. Clearly, the physician accounts for the lab's CT cutoff when determining the probability that a test patient is positive or negative.

In period 2, the physician's (and the lab's) goal is to set a CT cutoff (κ_{p2}) that minimizes the patient's expected cost of an erroneous diagnosis:

$$\xi(\kappa_{p2}) = \left(\frac{\alpha\phi_2}{D_2}\right) c_{FN}[1-G(\kappa_{p2})] + \left(1-\frac{\alpha\phi_2}{D_2}\right) c_{FP}H(\kappa_{p2}). \tag{24}$$

The solution of the first-order condition with respect to κ_{p2} yields

$$\kappa_{p2} = \frac{\left(\frac{\alpha\phi_2}{D_2}\right) c_{FN}}{\left(\frac{\alpha\phi_2}{D_2}\right) c_{FN} + \left(1-\frac{\alpha\phi_2}{D_2}\right) c_{FP}}. \tag{25}$$

For ease of exposition, we define

$$r \triangleq r_0 - \min_{\kappa_{p2}} \left\{ \left(\frac{\alpha\phi_2}{D_2}\right) c_{FN}[1-G(\kappa_{p2})] + \left(1-\frac{\alpha\phi_2}{D_2}\right) c_{FP}H(\kappa_{p2}) \right\}. \tag{26}$$

In period 1, the lab sets κ_l and C_2 to maximize its objective function:

$$\begin{aligned}\pi(\kappa_l, C_2) = {} & \mathbb{E}[r \cdot \min\{C_2, z_1(\kappa_l) + z_2(\kappa_l)R_0\}] - \gamma \cdot (C_2^2 - C_1^2) \\ & - C_1 \cdot \left\{ \frac{\alpha\phi}{\alpha\phi + \beta(1-\phi)} c_{FN}[1-G(\kappa_l)] + \frac{\beta(1-\phi)}{\alpha\phi + \beta(1-\phi)} c_{FP}H(\kappa_l) \right\},\end{aligned} \tag{27}$$

where the first term represents the lab's total expected revenue less patients' expected cost of erroneous diagnosis during period 2, the second term is the cost of building capacity, and the third term captures patients' expected cost of erroneous diagnosis during period 1. (The rest of the analysis is presented in the Appendix.)

We find the lab's period-2 capacity increases in κ_l and that $\kappa_l > \kappa_{p1}$ when ρ is sufficiently large. (As expected, the threshold depends on ϕ, α, β, $\bar{R}_0$, c_{FP}, and c_{FN}.) Since C_2 is increasing in κ_l, when the lab sets the CT cutoff, the corresponding period-2 capacity is higher than the capacity corresponding to a CT cutoff of κ_{p1}. If the lab reports the CT

value, the physician uses the cutoffs κ_{p1} in period 1 and κ_{p2} in period 2 to diagnose the patient. In period 2, the lab is indifferent between reporting and not reporting the CT value. However, similar to the main analysis, in period 1, the lab prefers not to disclose the CT value and to diagnose the patient using its own CT cutoff instead of disclosing the CT value to the physician. Since $\kappa_l > \kappa_{p1}$ and $\xi(\kappa)$ is increasing in κ for all $\kappa > \kappa_{p1}$, $\xi(\kappa_l) > \xi(\kappa_{p1})$. Furthermore, we find that the results presented in Proposition 5 hold qualitatively.

5.2 A Non-profit Lab without Capacity Constraint

Motivated by the observation that major diagnostic testing labs in the U.S. (e.g., Labcorp, Quest, etc.) are publicly traded private companies, we assume in the main model that the lab is concerned not only with patient utility but also with testing revenue. However, some labs (e.g., government and university labs) may not be concerned with test revenue. In this section, we look at a scenario in which a lab is concerned with the patient and society but not with its own profitability. Furthermore, to focus on non-capacity-based incentives, we assume the lab's fixed capacity is sufficient to meet testing demand. While acknowledging the possibility of undertesting during high demand periods, this model will focus on the issue of overdiagnosis. Finally, unlike in the main model, where we assumed the policymaker does not interfere with medical practice (as is the case in the U.S.), we allow the policymaker to do so in this section by specifying a CT cutoff.

Now we describe specifics of the model. The cost of false-positive diagnosis c_{FP} is composed of two parts: $c_{FP} = c_{FP}^P + c_{FP}^S$, where c_{FP}^P is the cost imposed on the patient and c_{FP}^S is the cost imposed on society. Similarly, the cost of false-negative diagnosis $c_{FN} = c_{FN}^P + c_{FN}^S$, where c_{FN}^P and c_{FN}^S are costs imposed on the patient and society, respectively. The lab, the physician, and the patient may all place different weights (w) on the cost imposed on the patient. (They put a weight $1 - w$ on the cost imposed on society.) For example, the patient may worry only about costs imposed on her, the physician may primarily care about her patient, and the lab may worry about society to a larger extent. The lab has no capacity constraint; therefore, all of the patients seeking testing get tested. The lab or the physician has no profit motives. We can write the sum of the weighted expected

cost of false-negative and false-positive diagnosis by

$$\xi(\kappa) = [1 - G(\kappa)]\left[w\, c_{FN}^{P} + (1-w)\, c_{FN}^{S}\right] + H(\kappa)\left[w\, c_{FP}^{P} + (1-w)\, c_{FP}^{S}\right]. \tag{28}$$

The policymaker cares about reducing the number of infections in period 2 in addition to lowering the cost imposed by incorrect diagnosis. The policymaker minimizes

$$\pi_{G}^{'} = \phi R_0 \left[1 - \alpha\left(2\kappa - \kappa^2\right)\right] + \xi(\kappa), \tag{29}$$

where the first term captures the mass of infectious patients at the start of period 2 given that the lab faces no capacity limitation, and the second term captures the cost of erroneous diagnosis. The policymaker moves first and sets the CT cutoff (κ_G) and decides whether to mandate CT-value reporting. The lab follows the CT cutoff (κ_G) to diagnose patients and reports the CT value as mandated by the policymaker. If CT-value reporting is not mandated by the policymaker, the lab decides whether to report the CT value voluntarily. If the lab reports the CT value, the physician ignores the lab's diagnosis and uses her own CT cutoff (κ_P) to diagnose the patient. However, if the lab does not report the CT value, the physician follows the lab's diagnosis. Other assumptions are the same as in the main model. The following proposition presents period-1 and period-2 CT-cutoff decisions.

Proposition 6. *If the lab faces no capacity constraint and cares only about costs that erroneous diagnosis impose on the patient and society,*

(i) in period 2, the lab, the physician, and the policymaker prefers to set the CT cutoff

$$\kappa_2 = \frac{w\, c_{FN}^{P} + (1-w)\, c_{FN}^{S}}{w\left(c_{FN}^{P} + c_{FP}^{P}\right) + (1-w)\left(c_{FN}^{S} + c_{FP}^{S}\right)}, \tag{30}$$

which is decreasing in w for

$$c_{FN}^{S}\, c_{FP}^{P} > c_{FP}^{S}\, c_{FN}^{P}.$$

(ii) If the policymaker can mandate a CT cutoff, in period 1, the policymaker mandates a

CT cutoff

$$\kappa_G = \frac{\phi R_0 \alpha + w\, c_{FN}^P + (1-w)\, c_{FN}^S}{\phi R_0 \alpha + w\left(c_{FN}^P + c_{FP}^P\right) + (1-w)\left(c_{FN}^S + c_{FP}^S\right)}, \tag{31}$$

which is greater than κ_2.

In the terminal period 2, similar to the lab and the physician, the policymaker is also concerned only about minimizing $\xi(\kappa)$. Therefore, all three prefer to use the same CT cutoff κ_2. The condition that ensures κ_2 is decreasing in w is intuitive. A lower CT cutoff would result in lower false-positive and higher false-negative rates. Therefore, a player who cares more about the patient would set a lower CT cutoff if a false-positive diagnosis imposes a large cost on the patient but a relatively smaller cost on society, whereas a false-negative diagnosis imposes a large cost on society relative to the cost it imposes on the individual. As expected, we observe that physicians, who primarily care about patients, advocate for smaller CT cutoffs compared with labs and policymakers, who care about society as well.

The CT cutoff is higher when it is set by the policymaker. The reason is that the policymaker is concerned not only with the cost associated with diagnoses, but also with the spread of the disease. The policymaker does not mandate CT-value reporting. The CT-value allows the physician (and other stakeholders who are primarily concerned with the patient) to disregard the κ_G-based diagnosis and instead use their own CT cutoff to arrive at a potentially different diagnosis. The ability of the physician to determine the patient's diagnosis interferes with the policymaker's goal of reducing disease spread in the population. Therefore, the policymaker does not require CT-value reporting. The lab does not voluntarily report CT values, because the lab is more concerned with society than the physician, who is primarily concerned with her patient. If the lab reports a CT value, the physician will disregard the lab's diagnosis and use a CT cutoff that is lower than both the policymaker's and the lab's desired cutoff to diagnose the patient. As a result, the lab does not report CT values in the first period. In the second period, however, the policymaker either lowers the CT cutoff or mandates CT-value reporting. In the absence of such a policymaker's action, the lab may choose to voluntarily report the CT value on the test result.

6. Concluding Remarks

The main contribution of this paper is to shed light on the interaction between diagnostic decision-making and capacity building when diagnoses have implications for future market demand for testing. When the lab provides diagnoses unaccompanied by fine-grained viral-load information, positive diagnoses with nearly non-existent viral loads (i.e., "overdiagnosis") may trigger contact-tracing efforts and hence boost market demand for testing. At the same time, such positive diagnoses lead to the quarantining of potentially infectious individuals on a wider scale, and thus help decelerate disease transmission and dampen future organic demand for testing. By analyzing this tradeoff, we find scenarios exist in which the lab has an incentive to engage in overdiagnosis to generate higher diagnosis-driven future demand. Overdiagnosis leads to a higher testing capacity, which is nevertheless insufficient to absorb the inflated demand, so undertesting arises along with overdiagnosis.

Beyond this insight, our model has implications for a policymaker's decision regarding whether to require labs to disclose viral loads on their test results. When the lab is mandated to disclose the viral-load information, it essentially forfeits its decision rights—in terms of the interpretation of the viral load, which may trigger subsequent interventions that shape future demand for testing—to physicians and other stakeholders. We show that although the concept of information disclosure is broadly appealing, when the policymaker intends to induce the lab to build a high testing capacity and reduce viral transmission, *not* mandating such disclosure may be in the best interest of the policymaker. Although the lab and the policymaker have divergent objectives, a CT cutoff that is higher than what is ideal for an individual patient is desirable to both.

In sum, we have developed a theoretical model that sheds light on two important issues (overdiagnosis and undertesting) exposed by the pandemic. By incorporating and connecting two key aspects of diagnostic testing for an infectious disease, namely, diagnostic threshold and testing capacity, we contribute to the ongoing debates about viral-load reporting (Mandavilli 2020b) and provide novel policy insights for the current and future pandemics. Our research calls for further analyses with additional considerations (e.g., the competition between multiple labs) and empirical investigations.

Our theory does not capture all the forces driving a lab's CT cutoff and capacity decisions. For example, a lab may set a high CT cutoff if it is concerned about potential liability arising from a false-negative diagnosis. The lab may also set a conservative high CT cutoff if viral loads in tested patients are expected to increase. There may be several other plausible reasons for a lack of CT reporting. For example, CT reporting may eliminate confusion in some cases and create confusion in others. Thus, the lack of CT reporting may be due to avoidance of confusion. Another reason may be that different entities may choose different cutoffs and diagnose the patient differently if CT values are made available.

Our work has several limitations that represent promising areas for future research. First, in practice, unethical behavior can result from organizational incentives, which can dilute an individual manager's moral responsibility. Note our model does not explicitly distinguish between organizational incentives and individual morality. Thus, the lab's decision to choose a high CT cutoff should not be interpreted as widespread immorality or unethical behavior among lab managers. Examining the interaction between organizational and individual incentives is a promising area for future research. Organizational incentives for overdiagnosis and undertesting may be either mitigated or reinforced by individual incentives, depending on the compensation structure of lab managers. Second, our model is agnostic about the source of the contact-tracing demand and we do not consider the interaction between organic (i.e., from community spread) and induced (i.e., from contact tracing) demand, which constitutes another venue for future investigation. For example, one can argue wider community spread can make contact tracing difficult and thus reduce the intensity of contact tracing. Third, in accordance with standard practice, we assume the price of testing is exogenous. Future research may involve relaxing this assumption and examining its impact on the lab's testing threshold and capacity decisions. The lab could conceivably increase its service fee to dampen demand and thus reduce the need to inflate the diagnostic criterion. Finally, pending data availability, our theoretical investigation offers opportunities for empirical testing to determine whether heterogeneous labs behave differently in ways that are consistent with our predictions.

Appendices

A1 Technical Proofs

Proof of Lemma 1. First, the first-order derivative of $z_1(\kappa_l)$ is

$$z_1'(\kappa_l) = 2\rho \cdot C_1 \cdot \frac{\alpha\phi\,(1-\kappa_l) + (1-\phi)\,\beta\kappa_l}{\alpha\phi + (1-\phi)\,\beta} > 0. \tag{A1}$$

Thus, $z_1(\kappa_l)$ increases in κ_l. Second, the first-order derivative of $z_2(\kappa_l)$ is

$$z_2'(\kappa_l) = -2(\alpha-\beta) \cdot C_1 \cdot (1-\kappa_l) \cdot \frac{\alpha\phi}{\alpha\phi + (1-\phi)\,\beta} < 0. \tag{A2}$$

Thus, $z_2(\kappa_l)$ decreases in κ_l. *Q.E.D.*

Proof of Proposition 1. Given κ_l, the first-order derivative of the lab's objective function with respect to C_2 can be reorganized, by Leibniz's rule, as

$$\begin{aligned}
\frac{\partial \pi(\kappa_l, C_2)}{\partial C_2} &= r \cdot \frac{\partial \mathbb{E}\min\{C_2, z_1(\kappa_l) + z_2(\kappa_l)\cdot R_0\}}{\partial C_2} - 2\gamma C_2 && \text{(A3)}\\
&= r \cdot \frac{\partial\left\{\int_0^{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}} [z_1(\kappa_l) + z_2(\kappa_l)\cdot R_0]dF(R_0) + \int_{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}}^{\bar{R}_0} C_2 dF(R_0)\right\}}{\partial C_2} - 2\gamma C_2 && \text{(A4)}\\
&= r \cdot \left[\frac{C_2}{z_2(\kappa_l)} - \frac{C_2}{z_2(\kappa_l)} + 1 - F\left(\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l)}\right)\right] - 2\gamma C_2 && \text{(A5)}\\
&= r \cdot \left[1 - F\left(\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l)}\right)\right] - 2\gamma C_2. && \text{(A6)}
\end{aligned}$$

Using the first-order condition gives

$$F\left(\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l)}\right) = 1 - \frac{2\gamma C_2}{r} \qquad \text{at } C_2 = C_2^*(\kappa_l). \tag{A7}$$

In addition, we can verify that the second-order condition is satisfied:

$$\frac{\partial^2 \pi(\kappa_l, C_2)}{\partial C_2^2} = -\frac{r}{z_2(\kappa_l)} f\left(\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l)}\right) - 2\gamma < 0, \tag{A8}$$

where $f(\cdot)$ is the probability density function (pdf) of R_0. Thus, the lab's objective function is concave in C_2. Then, using the assumption that R_0 follows a uniform distribution, the first-order condition simplifies to (16).

We have from (16) that $C_2^*(\kappa_l)$ increases in κ_l if $z_1(\kappa_l) + \bar{R}_0 \cdot z_2(\kappa_l)$ increases in κ_l. The first-order derivative of $z_1(\kappa_l) + \bar{R}_0 \cdot z_2(\kappa_l)$ is

$$z_1'(\kappa_l) + \bar{R}_0 \cdot z_2'(\kappa_l) = \frac{2C_1}{\beta + (\alpha - \beta)\phi} \cdot \{\{\phi\alpha[(\alpha - \beta)\bar{R}_0 - \rho] + (1 - \phi)\beta\rho\}\kappa_l - \phi\alpha[(\alpha - \beta)\bar{R}_0 - \rho]\}. \tag{A9}$$

The derivative $z_1'(\kappa_l) + \bar{R}_0 \cdot z_2'(\kappa_l) \geq 0$ if and only if

$$\{\phi\alpha[(\alpha - \beta)\bar{R}_0 - \rho] + (1 - \phi)\beta\rho\}\kappa_l - \phi\alpha[(\alpha - \beta)\bar{R}_0 - \rho] \geq 0, \tag{A10}$$

which gives

$$\rho \geq \frac{\phi\alpha(\alpha - \beta)\bar{R}_0 (1 - \kappa_l)}{\phi\alpha (1 - \kappa_l) + (1 - \phi)\beta\kappa_l} \tag{A11}$$

and completes the proof. *Q.E.D.*

Proof of Lemma 2. In this case, the lab discloses the CT value on the test result and the physician uses the CT-value information to diagnose the patient using her own CT cutoff. The physician chooses a CT cutoff of κ_p that minimizes $\xi(\kappa_p)$ as given in (2). Therefore, the physician sets $\kappa_p = c_{FN}/(c_{FN} + c_{FP})$. Substituting $\kappa_l = \kappa_p$ in (16) gives (18). *Q.E.D.*

Proof of Proposition 2. Equation (19) follows from Proposition 1. To prove (20), note the first-order partial derivative of the lab's objective function (15) in terms of κ_l is

$$r \cdot \frac{\partial \mathbb{E} \min\{C_2, z_1(\kappa_l) + z_2(\kappa_l) \cdot R_0\}}{\partial \kappa_l} - C_1 \cdot \{c_{FP} \cdot (2\kappa_l) + c_{FN} \cdot [-2(1 - \kappa_l)]\}. \tag{A12}$$

Using the Leibniz integral rule, we have

$$\frac{\partial \mathbb{E} \min\{C_2, z_1(\kappa_l) + z_2(\kappa_l) \cdot R_0\}}{\partial \kappa_l} \tag{A13}$$

$$= \frac{\partial \left\{ \int_0^{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}} [z_1(\kappa_l) + z_2(\kappa_l) \cdot R_0] dF(R_0) + \int_{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}}^{\bar{R}_0} C_2 dF(R_0) \right\}}{\partial \kappa_l} \tag{A14}$$

$$= C_2 \cdot \frac{\partial \frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}}{\partial \kappa_l} - C_2 \cdot \frac{\partial \frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}}{\partial \kappa_l} + \int_0^{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}} [z_1'(\kappa_l) + z_2'(\kappa_l) R_0] dF(R_0) \tag{A15}$$

$$= \int_0^{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}} [z_1'(\kappa_l) + z_2'(\kappa_l) R_0] dF(R_0). \tag{A16}$$

The first-order condition, along with (19), gives (20). *Q.E.D.*

PROOF OF LEMMA 3. If the physician follows the lab diagnosis, the patient's expected cost is

$$\xi(\kappa_l^*) = c_{FN}\left(1-\kappa_l^*\right)^2 + c_{FP}(\kappa_l^*)^2. \tag{A17}$$

However, if the physician discards the lab's diagnosis (and recommends quarantine or no quarantine), the patient's expected cost is $\min\{c_{FN}, c_{FP}\}$. Thus, the lab will not set a CT cutoff κ_l^* such that the physician ignores the lab's diagnosis. Equivalently,

$$c_{FN}\left(1-\kappa_l^*\right)^2 + c_{FP}(\kappa_l^*)^2 \leq \min\{c_{FN}, c_{FP}\}, \tag{A18}$$

which completes the proof. *Q.E.D.*

PROOF OF PROPOSITION 3. To prove Proposition 3, first, we establish the following claim:

Claim 1. *If* $\rho \geq (\alpha-\beta)\bar{R}_0$, $z_1'(\kappa_l) + z_2'(\kappa_l)\bar{R}_0 \geq 0$ *and* $C_2^*(\kappa_l)$ *increases in* κ_l *for* $0 < \kappa_l < 1$.

From the proof of Proposition 1, we have that $z_1'(\kappa_l) + z_2'(\kappa_l)\bar{R}_0 \geq 0$ if and only if

$$\rho \geq \frac{\phi\alpha(\alpha-\beta)\bar{R}_0\left(1-\kappa_l\right)}{\phi\alpha\left(1-\kappa_l\right) + (1-\phi)\beta\kappa_l}. \tag{A19}$$

In addition,

$$\frac{\phi\alpha(\alpha-\beta)\bar{R}_0\left(1-\kappa_l\right)}{\phi\alpha\left(1-\kappa_l\right) + (1-\phi)\beta\kappa_l} \leq (\alpha-\beta)\bar{R}_0 \quad \text{for all } 0 \leq \kappa_l \leq 1. \tag{A20}$$

Therefore, $z_1'(\kappa_l) + z_2'(\kappa_l)\bar{R}_0 \geq 0$, if $\rho \geq (\alpha-\beta)\bar{R}_0$.
Next, note from Proposition 1 that

$$C_2^*(\kappa_l) = \frac{r[z_1(\kappa_l) + z_2(\kappa_l)\bar{R}_0]}{r + 2z_2(\kappa_l)\bar{R}_0\gamma}. \tag{A21}$$

Under $\rho \geq (\alpha-\beta)\bar{R}_0$, $z_1'(\kappa_l) + z_2'(\kappa_l)\bar{R}_0 > 0$, so the numerator of the right-hand side of the above equation increases in κ_l. In addition, we know from Lemma 1 that its denominator always decreases in κ_l. Thus, if $\rho \geq (\alpha-\beta)\bar{R}_0$, $C_2^*(\kappa_l)$ increases in κ_l. Having established Claim 1, we now proceed to prove Proposition 3.

(i) From Lemma 1, $z_2'(\kappa_l) < 0$. Thus, for any realization of $R_0 \in [0, \bar{R}_0)$, we have from Claim 1

that,

$$z_1'(\kappa_l) + z_2'(\kappa_l)R_0 > z_1'(\kappa_l) + z_2'(\kappa_l)\bar{R}_0 > 0 \tag{A22}$$

if $\rho \geq (\alpha - \beta)\bar{R}_0$. Hence, the left-hand side of equation (20) must be strictly positive. Then, equation (20) gives

$$c_{FP} \cdot \kappa_l^* - c_{FN} \cdot (1 - \kappa_l^*) > 0. \tag{A23}$$

By contrast, κ_p satisfies

$$c_{FP} \cdot \kappa_p - c_{FN} \cdot (1 - \kappa_p) = 0. \tag{A24}$$

Because $c_{FP} \cdot H(\kappa_l) + c_{FN} \cdot [1 - G(\kappa_l)]$ increases in κ_l, comparing equations (A23) and (A24) gives $\kappa_l^* > \kappa_p$.

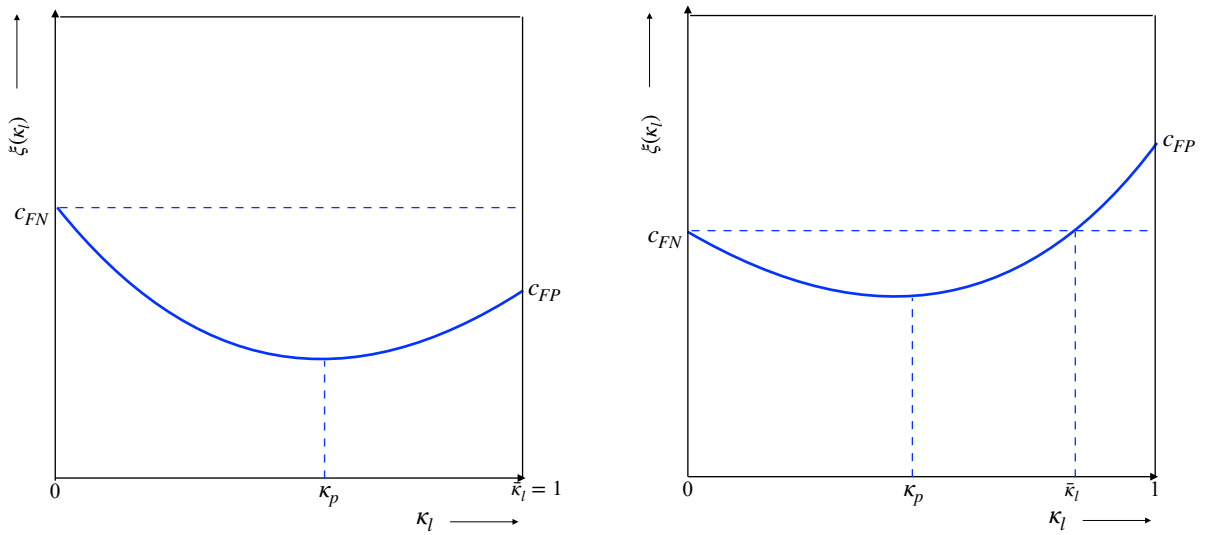
The lab's CT cutoff κ_l^* also satisfies (21), which imposes bounds on the value of κ_l^*, to ensure the physician follows the lab's diagnosis in the absence of CT-value information. Note that the lower bound of κ_l^* as specified by (21) is smaller than κ_p. This lower bound is redundant because we have established that $\kappa_l^* > \kappa_p$. Next, we examine the upper bound of κ_l^*. The lab will choose a CT cutoff κ_l^* that is below some $\overline{\kappa}_l$ such that the physician will not dismiss its binary diagnosis.

Note from Figure A1(b) that if $c_{FN} < c_{FP}$, $\overline{\kappa}_l < 1$. The physician will follow the lab's diagnosis for all κ_l^* satisfying $\kappa_p < \kappa_l^* < \overline{\kappa}_l$. The physician will be indifferent at $\kappa_l^* = \overline{\kappa}_l$. For most infectious diseases, $c_{FN} \geq c_{FP}$; that is, the cost of a false-negative diagnosis is higher than that of a false-positive diagnosis. If $c_{FN} \geq c_{FP}$, $\overline{\kappa}_l = 1$ (see Figure A1(a)). In other words, a physician would follow the lab's diagnosis for all $\kappa_p < \kappa_l^* < 1$ and be indifferent between following and not following the lab's diagnosis if $\kappa_l^* = 1$.

Therefore, if the lab sets the CT cutoff, it is higher than κ_p.

(ii) From Proposition 1 and Claim 1, $\frac{\partial C_2^*(\kappa_l)}{\partial \kappa_l} > 0$ for $\rho \geq (\alpha - \beta)\bar{R}_0$. In addition, from part (i) above, $\kappa_l^* > \kappa_p$. Therefore, $C_2^* = C_2^*(\kappa_l^*) > C_2^*(\kappa_p)$.

(iii) If the lab sets its own CT cutoff and testing capacity, the value of its objective function is $\pi\left(\kappa_l^*, C_2^*\left(\kappa_l^*\right)\right)$. However, if the lab reports the viral load (CT value) on the test result, the



(a) The case in which $c_{FN} \geq c_{FP}$.

(b) The case in which $c_{FN} < c_{FP}$.

Figure A1: Illustration of the Effective Range of the Lab's Choice of κ_l

physician sets the CT cutoff at κ_p. In this case, the lab sets capacity at $C_2^*(\kappa_p)$. Because the lab's objective function $\pi(\kappa_l, C_2)$ given in equation (15) is maximized at κ_l^* and $C_2^*(\kappa_l^*)$, it must be that $\pi(\kappa_l^*, C_2^*(\kappa_l^*)) > \pi(\kappa_p, C_2^*(\kappa_p))$. Therefore, the lab prefers setting κ_l itself and diagnosing the patient to reporting the CT value on the test result. *Q.E.D.*

Proof of Proposition 4

(i) Substituting the expressions of $G(\kappa_p)$ and $H(\kappa_p)$ in (2) gives,

$$\xi(\kappa_p) = c_{FN}\left(1 - \kappa_p^2\right) + c_{FP}\kappa_p^2.$$

It is straightforward that $\xi(\kappa_p)$ is minimized at $\kappa_p = \frac{c_{FN}}{c_{FN}+c_{FP}}$. The total expected cost $\xi(\kappa_p)$ at this cutoff is $\frac{c_{FN}c_{FP}(c_{FP}+3c_{FN})}{(c_{FN}+c_{FP})^2}$. Equivalently, the patient welfare is maximized at $\kappa_p = \frac{c_{FN}}{c_{FN}+c_{FP}}$. From Proposition 3(i), in period 1, the lab sets $\kappa_l^* > \kappa_p$. Since $\frac{\partial \xi(\kappa)}{\partial \kappa} > 0$ for $\kappa > \kappa_p$, $\xi(\kappa_l^*) > \xi(\kappa_p)$. Equivalently, patient welfare is lower when the patient is diagnosed using a CT cutoff of κ_l^* than when using a CT cutoff of κ_p.

(ii) First, we establish how κ_l^* responds to changes in parameters γ, r, ρ, and C_1. For ease of presentation, denote the integrand $\frac{r}{2\tilde{R}_0 C_1}\left(z_1'(\kappa_l) + z_2'(\kappa_l)R_0\right)$ on the left-hand side of (20) by $I(R_0)$ and the upper bound $\frac{\tilde{R}_0[r+2\gamma(C_1 - z_1(\kappa_l))]}{r+2\gamma\tilde{R}_0 z_2(\kappa_l)}$ by b. Substituting $z_1(\kappa_l)$ and $z_2(\kappa_l)$ from (12)–(13) and simplifying gives

$$I\left(R_0\right)=\frac{r}{\bar{R}_0}\left(\frac{\alpha\phi\left(1-\kappa_l\right)\left[\rho-\left(\alpha-\beta\right)R_0\right]+\rho\left(1-\phi\right)\beta\kappa_l}{\alpha\phi+\left(1-\phi\right)\beta}\right). \tag{A25}$$

An examination of the partial derivatives of $I\left(R_0\right)$ in equation (A25) with respect to parameters r, γ, ρ, and C_1 give $\frac{\partial I(R_0)}{\partial r}>0$, $\frac{\partial I(R_0)}{\partial \gamma}=0$, $\frac{\partial I(R_0)}{\partial \rho}>0$, and $\frac{\partial I(R_0)}{\partial C_1}=0$. Similarly, the partial derivatives of the upper bound b give $\frac{\partial b}{\partial r}>0$, $\frac{\partial b}{\partial \gamma}<0$, $\frac{\partial b}{\partial \rho}>0$, and $\frac{\partial b}{\partial C_1}>0$.

- $\frac{\partial I(R_0)}{\partial r}>0$ and $\frac{\partial b}{\partial r}>0$ mean the left-hand side of (20) is increasing in r. In addition, the right-hand side of (20) is independent of r and increasing in κ_l^*, which gives $\frac{\partial \kappa_l^*}{\partial r}>0$.
- $\frac{\partial I(R_0)}{\partial \gamma}=0$ and $\frac{\partial b}{\partial \gamma}<0$ mean the left-hand side of (20) is decreasing in γ, which gives $\frac{\partial \kappa_l^*}{\partial \gamma}<0$.
- $\frac{\partial I(R_0)}{\partial \rho}>0$ and $\frac{\partial b}{\partial \rho}>0$ mean the left-hand side of (20) is increasing in ρ, which gives $\frac{\partial \kappa_l^*}{\partial \rho}>0$.
- $\frac{\partial I(R_0)}{\partial C_1}=0$ and $\frac{\partial b}{\partial C_1}>0$ mean the left-hand side of (20) is increasing in C_1, which gives $\frac{\partial \kappa_l^*}{\partial C_1}>0$.

From part (i) of the proposition, a higher κ_l^* results in lower patient welfare. Therefore, patient welfare (corresponding to the lab's period-1 CT cutoff κ_l^*) is increasing in γ and decreasing in r, ρ, and C_1. *Q.E.D.*

Proof of Proposition 5. In the case in which viral-load reporting is mandatory, the lab reports the CT value as part of the test result, and the physician sets the CT cutoff at κ_p (in both periods 1 and 2). In this case, the lab sets period-2 capacity at $C_2^*(\kappa_p)$. However, in the absence of such mandatory-reporting requirements, the lab sets the period-1 CT cutoff at κ_l^* and period-2 capacity at $C_2^*\left(\kappa_l^*\right)$. Because period 2 is the terminal period, the period-2 CT cutoff is set at κ_p regardless of whether the lab or the physician sets it. Therefore, the comparison of mandatory and voluntary viral-load reporting is equivalent to the comparison of (1) period-1 CT cutoff κ_p and period-2 capacity $C_2^*(\kappa_p)$, and (2) period-1 CT cutoff κ_l^* and period-2 capacity $C_2^*(\kappa_l^*)$, respectively.

(i) Because the patient's expected cost from erroneous diagnosis, as defined in (2), is minimized (equivalently, patient utility is maximized) at $\kappa=\kappa_p$, it is straightforward that a mandatory CT-value reporting requirement results in a higher expected utility of the tested patient than in the absence of a mandatory-reporting requirement.

(ii) The mass of newly infectious patients at the beginning of period 2 depends on the period-1 CT cutoff (κ) and is given by $\phi_2(\kappa) = \left[\phi - \frac{\alpha\phi}{\alpha\phi+(1-\phi)\beta}(2\kappa - \kappa^2)C_1\right]R_0$. Note $\frac{\partial\phi_2}{\partial\kappa} < 0$ (i.e., a higher CT cutoff results in fewer infections). $\kappa_l^* > \kappa_p$ gives $\phi_2(\kappa_l^*) < \phi_2(\kappa_p)$. That is, a mandatory viral-load reporting requirement results in a more widespread infection in the subsequent period.

(iii) Finally, note that under a CT cutoff of κ_l, the lab's optimal service level is

$$\Pr(z_1(\kappa_l) + z_2(\kappa_l)R_0 \le C_2) = F\left(\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}\right), \tag{A26}$$

which, according to the proof of Proposition 1, is equal to

$$1 - \frac{2\gamma[C_2^*(\kappa_l) - C_1]}{r}. \tag{A27}$$

Because $\rho \ge (\alpha - \beta)\bar{R}_0\left(1 + \frac{2\gamma C_1}{r}\right)$, we have from Claim 1 that $C_2^*(\kappa_l)$ increases in κ_l. In addition, $\kappa_l^* > \kappa_p$. Therefore, $C_2^*(\kappa_l^*) > C_2^*(\kappa_p)$, and the service level is lower corresponding to a CT cutoff of κ_l^* than κ_p. *Q.E.D.*

Prevalence-Dependent CT Cutoff (Section 5.1)

In period 1, the lab sets κ_l and C_2 to maximize its objective function in (27). The first term of the right-hand side of (27) is

$$\mathbb{E}[r\min\{C_2, z_1(\kappa_l) + z_2(\kappa_l)R_0\}] = \int_0^{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}} r[z_1(\kappa_l) + z_2(\kappa_l)R_0]dF(R_0) + \int_{\frac{C_2 - z_1(\kappa_l)}{z_2(\kappa_l))}}^{\bar{R}_0} rC_2 dF(R_0).$$

In the above equation, r, as defined by (26), depends on κ_l (because both $D_2 = z_1(\kappa_l) + z_2(\kappa_l)R_0$ and $\phi_2 = \{\phi - \alpha\phi(2\kappa_l - \kappa_l^2)C_1/[\alpha\phi + (1-\phi)\beta]\}R_0$ depend on κ_l); thus, the lab's period-1 problem (without any additional assumptions) is analytically intractable.

We solve the lab's problem with two different approaches.

In our first approach, we make the simplifying assumption that r is an exogenous variable. This assumption can be considered reasonable if r_0 is much larger than $\frac{c_{FN}c_{FP}}{(\sqrt{c_{FN}}+\sqrt{c_{FP}})^2}$, which is the maximum magnitude of the second term in the right-hand side of equation (26). For example, consider the scenario where $r_0 = 2$, $c_{FN} = 1$, and $c_{FP} = 0.2$. In this case, $\frac{c_{FN}c_{FP}}{(\sqrt{c_{FN}}+\sqrt{c_{FP}})^2} = 0.095$, a value considerably smaller than r_0. This condition is more likely to be satisfied if there is a larger difference between the values of c_{FN} and c_{FP}. Under this condition, the first-order partial derivative

of the lab's objective function (27) with respect to κ_l is

$$\frac{\partial \mathbb{E}[r \min\{C_2, z_1(\kappa_l) + z_2(\kappa_l) \cdot R_0\}]}{\partial \kappa_l} - C_1 \cdot \{c_{FP} \cdot \frac{\beta(1-\phi)}{\alpha\phi + \beta(1-\phi)} \cdot (2\kappa_l) + c_{FN} \cdot \frac{\alpha\phi}{\alpha\phi + \beta(1-\phi)} \cdot [-2(1-\kappa_l)]\}. \tag{A28}$$

We can solve the lab's period-1 problem analytically and obtain the following solution:

Proposition A1. *If the lab sets its own CT cutoff and the physician follows the lab's diagnosis, the lab's optimal capacity C_2^* is*

$$C_2^* = \frac{[z_1(\kappa_l^*) + z_2(\kappa_l^*)\bar{R}_0]r}{r + 2z_2(\kappa_l^*)\bar{R}_0\gamma},$$

where κ_l^* satisfies

$$\frac{r}{2\bar{R}_0 C_1} \int_0^{\frac{\bar{R}_0[r - 2\gamma z_1(\kappa_l^*)]}{r + 2\gamma \bar{R}_0 z_2(\kappa_l^*)}} [z_1'(\kappa_l^*) + z_2'(\kappa_l^*)R_0] dR_0 = c_{FP} \cdot \frac{\beta(1-\phi)}{\alpha\phi + \beta(1-\phi)} \cdot \kappa_l^* - c_{FN} \cdot \frac{\alpha\phi}{\alpha\phi + \beta(1-\phi)} \cdot (1 - \kappa_l^*).$$

Using the modified proposition above, we can show that all of our main results hold qualitatively. Because the proofs are similar to those for Propositions 1–5, we will not repeat them here.

In the second approach, we solve the lab's problem numerically. Consistent with our main results, we find the lab's period-2 capacity increases in κ_l and that $\kappa_l > \kappa_{p1}$ when ρ is sufficiently large. (As expected, the threshold depends on ϕ, α, β, $\bar{R}_0$, c_{FP}, and c_{FN}.) Here is an illustrative example with the following parameter values: $r_0 = 3, \gamma = 3.5, \alpha = 0.6, \beta = 0.2, \phi = 0.15, C_1 = 0.15, \rho = 2, c_{FP} = 0.5$, and $\bar{R}_0 = 3$. We vary c_{FN} from 0.5 to 2.0 in steps of 0.01. Figure A2 shows $\kappa_l > \kappa_{p1}$ in the parameter space that satisfies the condition on ρ. Since C_2 is increasing in κ_l, when the lab sets the CT cutoff, the corresponding period-2 capacity is higher than the capacity corresponding to a CT cutoff of κ_{p1}.

In the case in which the lab reports the CT value, the physician uses the cutoffs κ_{p1} in period 1 and κ_{p2} in period 2 to diagnose the patient. In period 2, the lab is indifferent between reporting and not reporting the CT value. However, similar to our main analysis, we find that in period 1, the lab prefers not to disclose the CT value and to diagnose the patient using its own CT cutoff rather than disclosing the CT value to the physician.

Since $\kappa_l > \kappa_{p1}$ and $\xi(\kappa)$ is increasing in κ for all $\kappa > \kappa_{p1}$, $\xi(\kappa_l) > \xi(\kappa_{p1})$. Furthermore, we find that the results presented in Proposition 5 hold qualitatively.

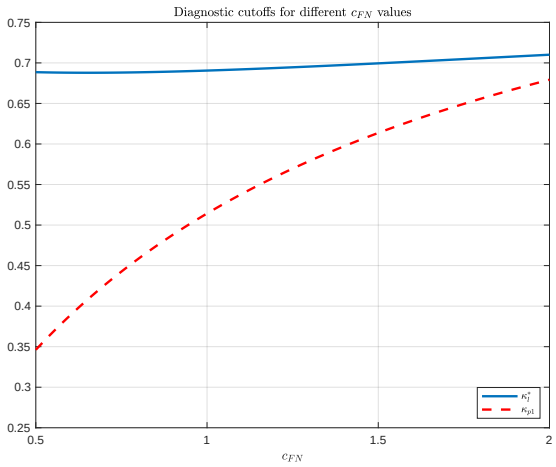


Figure A2: Comparison of κ_l^* and κ_{p1} for different values of c_{FN} under an alternative method for determining diagnostic cutoffs

Proof of Proposition 6

(i) Substituting the expressions of $G(\kappa)$ and $H(\kappa)$ in equation (28), we get

$$\xi(\kappa) = \left[1-\kappa^2\right]\left[w\, c_{FN}^P + (1-w)\, c_{FN}^S\right] + \kappa^2\left[w\, c_{FP}^P + (1-w)\, c_{FP}^S\right]. \tag{A29}$$

In period-2, which is the terminal period, the policymaker does not need to worry about the disease spread anymore. Therefore, incentives of all the players are aligned: in period 2, they prefer to set a threshold $\kappa = \kappa_2$ that mininizes $\xi(\kappa)$ given in equation (A29). Solving the first-order condition with respect to κ gives equation (30). The second-order derivative is positive, which confirms that $\xi(\kappa)$ is minimized at $\kappa = \kappa_2$. It is straightforward to show $\frac{\partial \kappa_2}{\partial w} < 0$ if

$$c_{FN}^S\, c_{FP}^P > c_{FP}^S\, c_{FN}^P.$$

(ii) In period 1, the policymaker sets a CT cutoff κ_G that minimizes (29). Solving the first-order condition with respect to κ gives equation (31). The second-order condition is satisfied. It is straightforward that $\kappa_G > \kappa_2$. *Q.E.D.*

Funding and Competing Interests

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

References

Adida E, Dai T (2023) Impact of physician payment scheme on diagnostic effort and testing. *Management Sci.* Forthcoming.

Arora A, Fosfuri A (2005) Pricing diagnostic information. *Management Sci.* 51(7):1092–1100.

Arrow KJ (1963) Uncertainty and the welfare economics of medical care. *Amer. Economic Rev.* 53(5):941–973.

Becker GS (2020) Some economics of flu pandemics. *Chicago Booth Rev.* (Summer) pp. 54–55. https://www.chicagobooth.edu/-/media/project/chicago-booth/chicago-booth-review/past-issue-pdfs/20202junelowres.pdf. Originally published in 2009 on *The Becker-Posner Blog* under the title of "The economics of the flu epidemic." http://bit.ly/Becker09blog. Both Accessed January 1, 2022.

Bentley TG, Catanzaro A, Ganiats TG (2012) Implications of the impact of prevalence on test thresholds and outcomes: lessons from tuberculosis. *BMC Res. Notes* 5(1):563–569.

Bhardwaj P (2001) Delegating pricing decisions. *Marketing Sci.* 20(2):143–169.

Bhardwaj P, Chen Y, Godes D (2008) Buyer-initiated vs. seller-initiated information revelation. *Management Sci.* 54(6):1104–1114.

Brodersen J, Schwartz LM, Heneghan C, O'Sullivan JW, Aronson JK, Woloshin S (2018) Overdiagnosis: What it is and what it isn't. *BMJ EBM* 23(1):1–3.

Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, Salanti G, Low N (2020) Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med.* 17(9):e1003346.

Caraguel CGB, Stryhn H, Gagné N, Dohoo IR, Hammell KL (2011) Selection of a cutoff value for real-time polymerase chain reaction results to fit a diagnostic purpose: Analytical and epidemiologic approaches. *J. Vet. Diagnostic Invest.* 23(1):2–15.

CDC (2020) COVID-19 pandemic planning scenarios. Centers for Disease Control and Prevention (September 10). http://bit.ly/cdcestimates. Accessed January 31, 2021.

Chen N, Hu M, Zhang C (2022) Capacitated SIR model with an application to COVID-19. *University of Toronto Working Paper.*

Cutler DM, Summers LH (2020) The COVID-19 pandemic and the $16 trillion virus. *JAMA* 324(15):1495–1496.

Dai T, Singh S (2020) Conspicuous by its absence: Diagnostic expert testing under uncertainty. *Marketing Sci.* 39(3):540–563.

Darby MR, Karni E (1973) Free competition and the optimal amount of fraud. *J. Law Econom.* 16(1):67–88.

de Véricourt F, Gurkan H, Wang S (2020) Informing the public about a pandemic. *Management Science* 27(10):6350–6357.

Dogan M, Jacquillat A, Yildirim P (2019) Strategic automation and decision-making authority. *MIT Sloan School of Management Working Paper.*

Doi, SAR (2013). Using and interpreting diagnostic tests with quantitative results. In S. A. R. Doi and G. M. Williams (Eds.), *Methods of Clinical Epidemiology* (pp. 67–78). Springer Berlin Heidelberg.

Dranove D (1988) Demand inducement and the physician/patient relationship. *Econom. Inquiry* 26(2):281–298.

Durbin E, Iyer G (2009) Corruptible advice. *Amer. Econom. J. Microeconom.* 1(2):220–242.

Dye RA (1985) Disclosure of nonproprietary information. *J. Accounting Res.* 23(1):123–145.

Ekelund S (2011) ROC curves — what are they and how are they used? *Acute Care Testing* (January). https://acutecaretesting.org/en/articles/roc-curves-what-are-they-and-how-are-they-used. Accessed June 7, 2023.

Eliashberg J, Steinberg R (1993) Chapter 18 Marketing-production joint decision-making. *Handbooks in Operations Research and Management Science* (Elsevier B.V., Amsterdam, Netherlands):827–880.

Ely, J, Galeotti A, Jann O, Steiner J (2021) Optimal Test Allocation. *Journal of Economic Theory* 193:105236.

Esserman LJ, Thompson IM, Reid B (2013) Overdiagnosis and overtreatment in cancer. *JAMA* 310(8):797–798.

Feehan DM, Mahmud AS (2021) Quantifying population contact patterns in the United States during the COVID-19 pandemic. *Nature Comm.* 12(1):893.

Foley J (2020) How U.S. made Covid-19 tests a profitable disaster. Reuters (July 23). https://reut.rs/2Lt0pwb. Accessed January 1, 2022.

Gal-Or E, Gal-Or M, Dukes A (2007) Optimal information revelation in procurement schemes. *RAND J. Econom.* 38(2):400–418.

Grossman SJ (1981) The informational role of warranties and private disclosure about product quality. *J. Law Econom.* 24(3):461–483.

Grynbaum MM (2020) NBC faces blowback for holding Trump's town hall opposite Biden's. *New York Times* (October 14). https://nyti.ms/31JhHKM. Accessed January 1, 2022.

Guo L (2020) Upstream exploitation and strategic disclosure. *Marketing Sci.* 39(5):923–938.

Guo L (2022a) The credibility of communication in a pandemic. *J. Marketing Res.* Forthcoming.

Guo L (2022b) Lockdown or quarantine: When should mandatory isolation be targeted? *Working Paper.*

Guo L, Xu W (2023) "We are the world": When more equality improves efficiency and antipandemic consumptions are intervened. *Marketing Sci.* 42(2):214–232.

Guo L, Zhao Y (2009) Voluntary quality disclosure and market interaction. *Marketing Sci.* 28(3):488–501.

Iyer G, Singh S (2018) Voluntary product safety certification. *Management Sci.* 64(2):695–714.

Iyer G, Singh S (2022) Persuasion contest: Disclosing own and rival information. *Marketing Sci.* 41(4):254–281.

Jain S, Jónasson JO, Pauphilet J, Ramdas K (2024) Robust combination testing: Methods and application to COVID-19 detection. *Management Sci.* 70(4):2661–2681.

Jiang B, Ni J, Srinivasan K (2014) Signaling through pricing by service providers with social preferences. *Marketing Sci.* 33(5):641–654.

Jovanovic B (1982) Truthful disclosure of information. *Bell J. Econom.* 13(1):36–44.

Lal R (1986) Delegating pricing responsibility to the salesforce. *Marketing Sci.* 5(2):159–168.

Madrigal AC, Meyer R (2020) Why Trump's rapid-testing plan worries scientists: Experts were already divided on the right way to deploy new coronavirus tests. Then the White House barged ahead. *Atlantic* (October 9). http://bit.ly/rapidcovidtesting. Accessed January 1, 2022.

Mak H, Dai T, Tang CS (2022) Managing two-dose COVID-19 vaccine rollouts with limited supply: Operations strategies for distributing time-sensitive resources. *Production Oper. Management* 31(12):4424–4442.

Mandavilli A (2020a) Your Coronavirus test is positive. Maybe it shouldn't be. *New York Times* (August 29). https://nyti.ms/31Vyv1B. Accessed January 1, 2022.

Mandavilli A (2020b) You're infected with the coronavirus. But how infected?. *New York Times* (December 29). https://nyti.ms/3bpcaPx. Accessed January 1, 2022.

Mayzlin D, Shin J (2011) Uninformative advertising as an invitation to search. *Marketing Sci.* 30(4):666–685.

McGuire TG (2000) Physician agency. Culyer AJ, Newhouse JP eds. *Handbook of Health Economics, Volume 1, Part A* (Elsevier, Amsterdam, The Netherlands), 461–536.

McNeil BJ, Keeler E, Adelstein SJ (1975) Primer on certain elements of medical decision making. *New England J. Med.* 293(5):211–215.

Milgrom PR (1981) Good news and bad news: Representation theorems and applications. *Bell J. Econom.* 12(2):380–391.

Mullainathan S, Obermeyer Z (2022) Diagnosing physician error: A machine learning approach to low-value health care. *Quart. J Econom.* 137(2):679–727.

Özer Ö, Subramanian U, Wang Y (2018) Information sharing, advice provision, or delegation: What leads to higher trust and trustworthiness? *Management Sci.* 64(1):474–493.

Rosenthal E (2021) My epic, infuriating quest for a covid test. *Washington Post* (December 27). https://wapo.st/3z4l0vy. Accessed January 1, 2022.

Rao S (2020) Covid-19 test reports must also state cycle threshold value: Doctors. *Times of India* (September 6). http://bit.ly/toictvalue. Accessed January 1, 2022.

Salis AD (2009) Applications in clinical microbiology. Logan J, Logan JMJ, Edwards KJ, Saunders NA, eds. *Real-Time PCR: Current Technology and Applications* (Caister Academic Press, Norfolk, United Kingdom), 177–220.

Sarvary M (2002) Temporal differentiation and the market for second opinions. *J. Marketing Res.* 39(1):129–136.

Service RF (2020) A call for diagnostic tests to report viral load. *Sci.* 370(6512):22.

Shear MD, Goodnough A, Kaplan S, Fink S, Thomas K, Weiland N (2020) The lost month: How a failure to test blinded the U.S. to Covid-19. *New York Times* (March 28). https://nyti.ms/3joYBj5. Accessed January 1, 2022.

Shirts BH (2020) We'll see more shortages of diagnostic tests if the FDA has its way. *STAT* (April 15). http://bit.ly/stat-testing. Accessed January 1, 2022.

Singh S (2017) Competition in corruptible markets. *Marketing Sci.* 36(3):361–381.

Sun M, Tyagi RK (2020) Product fit uncertainty and information provision in a distribution channel. *Production Oper. Management* 29(10):2381–2402.

TWiV (2020) TWiV 641: COVID-19 with Dr. Anthony Fauci. *This Week in Virology* (July 17). https://www.microbe.tv/twiv/twiv-641/. Accessed January 1, 2022.

US Census (2019) Historical households tables. United States Census Bureau (October 10). http://bit.ly/censushhsize. Accessed January 1, 2022.

Verrecchia RE (1983) Discretionary disclosure. *J. Accounting Econom.* 5(April):179–194.

Wang G, Zhang M, Dai T (2024) The spillover effect of suspending non-essential surgery: Evidence from kidney transplantation. *Johns Hopkins University Working Paper.*

Welch HG, Schwartz L, Woloshin S (2011) *Overdiagnosed: Making People Sick in the Pursuit of Health* (Beacon Press, Boston).

Wolinsky A (1993) Competition in a market for informed experts' services. *RAND J. Econom.* 24(3):380–398.

Wu KJ (2020) In coronavirus testing, false positives are more hazardous than they might seem. *New York Times* (October 25). https://nyti.ms/34zqRf1. Accessed January 1, 2022.

Zhang J (2014) Policy and inference: The case of product labeling. *MIT Sloan School of Management Working Paper.*